Caged phosphopeptides and phosphoproteins:
Probes to dissect the role of phosphorylation in complex signaling pathways

by

Elizabeth Maura Vogel

B.A. Chemistry
Hamilton College, 2001

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

at the

Massachusetts Institute of Technology

February 2007

Signature of Author: _____
Department of Chemistry
December 19, 2006

Certified by: _____
Barbara Imperiali
Class of 1922 Professor of Chemistry and Professor of Biology
Thesis Supervisor

Accepted by: _____
Robert W. Field
Haslam and Dewey Professor of Chemistry
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Professor Timothy F. Jamison_____
Chairman

Professor Barbara Imperiali_____
Thesis Supervisor

Professor Alice Y. Ting_____

Caged phosphopeptides and phosphoproteins:
Probes to dissect the role of phosphorylation in complex signaling pathways

by
Elizabeth Maura Vogel

Submitted to the Department of Chemistry
on December 19, 2006 in partial fulfillment of the requirements
of the Degree of Doctor of Philosophy in Organic Chemistry

ABSTRACT

Protein phosphorylation is a central regulatory mechanism in signal transduction pathways and cellular migration. Current genetic strategies for the study of phosphorylation, including gene knockout and point mutation, are limited in providing temporal information. As a complement to these techniques, the synthesis and semisynthesis of probes that enable researchers to observe the downstream effects of kinase-mediated phosphorylation in "real time" are presented in this thesis.

The release of a physiologically-relevant concentration of a phosphopeptide with temporal and spatial control is accomplished by the photolysis of a photolabile precursor, a caged phosphopeptide. The synthesis and application of N^{α}-Fmoc-protected 1-(2-nitrophenyl) ethyl (NPE) caged phosphothreonine, serine, and tyrosine building blocks facilitate the straightforward assembly of any caged phosphopeptide through Fmoc-based solid phase peptide synthesis. Removal of the NPE caging group by irradiation with long-wavelength UV light generates a concentration burst of the corresponding phosphopeptide.

In addition, the installation of a caged phosphoamino acid into a full-length, multi-domain protein, the cellular migration protein paxillin, is described. A strategy, which is applicable to any expressible protein target, is detailed for the semisynthesis of a paxillin variant with a caged phosphorylated tyrosine at residue 31 of the 557-residue protein using native chemical ligation. Paxillin is a 61-kDa protein known to orchestrate the interaction of signaling proteins involved in cell migration by acting as a molecular adaptor, with the creation of specific binding sites dependent on paxillin phosphorylation. Therefore, the semisynthetic probe comprises the entire paxillin macromolecule, including all other binding and localization domains, which are essential for creating a native-like system to probe the effect of phosphorylation. The comprehensive biochemical characterization of the paxillin probe and quantification of uncaging following irradiation with long-wavelength UV light are also described.

Additionally, the strategy developed for the paxillin semisynthesis was applied to incorporate a different unnatural amino acid, the fluorescent chemosensing residue Sox, into a protein-domain sensor for ERK2 kinase activity. The protein domain sensor demonstrated significantly improved sensitivity for ERK2 phosphorylation over the corresponding peptide probe.

Thesis Supervisor: Barbara Imperiali
Title: Class of 1922 Professor of Chemistry and Professor of Biology

Acknowledgements

First, I thank my advisor, Professor Barbara Imperiali, for her creativity, her wealth of ideas, and her support. She has created a lab environment that has been a joy to work in for the past five years. In addition, I have learned much from the breadth of her scientific knowledge and interests, and through her inspiring skills at teaching, both within the group and in the classroom. It has been a true privilege to be a part of the Imperiali lab.

Next, I thank all the past and present members of the Imperiali group, who have made day-to-day research particularly fun, and who have provided so many ideas, patient instruction, and great friendships. Debbie Rothman and Melissa Shults have been both appreciated scientific advisors and dear friends. I thank Debbie for taking me under her wing when I first joined the lab, and for her great energy and enthusiasm for all she does. To Melissa Shults, for inspiring me with her dedication and work ethic and for her thoughtful and sincere attention to others. Many thanks to Jebrell Glover for his expertise and patience in teaching me molecular biology. To Elvedin Lukovic, Dora Carrico Moniz, Mary O'Reilly, Bianca Sculimbrene, Anne Reynolds, Galen Loving, Langdon Martin and Eranthie Weerepana, I thank them for their trusted scientific advice and their valued friendships. I would also like to thank and specially acknowledge the Imperiali-Red Sox alliance of Debbie, Melissa, Bianca, Langdon, Anne, and Seungib Chou. A special thanks to Brenda, Dora, Anne, Alex, and Elvedin for each tackling a chapter of my thesis with their brilliant editing skills. Thank you and best of luck to Matthieu Sainlos, Andreas Aemissegger, Nelson Olivier, Mark Chen, and our great second years, Brenda Goguen, Meredith Hartley, Wendy Iskenderian, and Angelyn Larkin.

I would also like to thank Professor Tim Jamison for his guidance as my thesis chair. An enormous thank you to Elizabeth Fong and Susan Brighton for all that they do to make the lab and department run smoothly.

I have many people I would like to acknowledge outside of MIT. I thank Professor Ian Rosenstein for first sparking my enthusiasm for organic chemistry and for being and incredible teacher and mentor. To Chris Steed and John Doench, for being like brothers to me, and for their support and friendship. I especially thank my amazing girlfriends, Janet Damaske, Meg Hern Steed, my roommate Eliz Guancial, Lynne Myrth, Amy Lawrence, and Catherine Wolf.

I thank Alex Taylor for being my most trusted editor, for his fantastic ideas, and for all the happiness he brings me.

And most importantly, I thank my family (ALL of you), and particularly Alex, Mom, Dad, Sara, Kev, Ty, Jilly, and Auntie. You make my life in every way joyful. I dedicate this thesis to you.

Table of Contents

List of Figures

List of Schemes

List of Abbriviations

Standard one-letter and three-letter codes are used for the 20 natural amino acids.

Ac	acetyl
Acm	acetamidomethyl
ATP	adenosine-5'-triphosphate
BAL	backbone amide linker
BME	beta-mercaptoethanol
BSA	bovine serum albumin
calcd	calculated
CBD	chitin binding domain
CHEF	chelation enhanced fluorescence
Cit	citrulline
CNBr	cyanogen bromide
C-terminus	carboxy-terminus
DCM	dichloromethane
DDI	DNA directed immobilization
DIPEA	*N*,*N*-diisopropylethylamine
DMF	*N*,*N*-dimethylformamide
DMSO	dimethylsulfoxide
DTT	dithiothreitol
ECM	extracellular matrix
EDT	ethanedithiol
EDTA	ethylenediaminetetraacetic acid
ENLYFQC	TEV protease cleavage sequence (Glu-Asn-Leu-Tyr-Phe-Gln-Cys)
EPL	expressed protein ligation
EPR	electron paramagnetic resonance
ERK	extracellular signal regulated kinase
ESI-MS	electrospray ionization mass spectrometry
Φ	quantum yield of uncaging
FAK	focal adhesion kinase
Fmoc	*N*-α-fluorenylmethoxylcarbonyl
FHA	fork head associated
FRET	fluorescence resonance energy transfer
GFP	green fluorescent protein
GIT	GRK interactor 1
GPI	glycosyl phosphatidylinositol
GST	glutathione S-transferase
IMPACT	intein-mediated purification with affinity chitin binding tag
HATU	*O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium Hexafluorophosphate
Hepes	*N*-(2-hydroxyethyl)piperazine-*N*'-ethanesulfonic acid
HOBt	*N*-hydroxylbenzotriazole
HPCL	high pressure liquid chromatography

IEGR	FLAG protease cleavage sequence (Ile-Glu-Gly-Arg)
IPTG	isopropyl-β-D-thiogalactopyranoside
MALDI MS	matrix assisted laser desorption ionization mass spectrometry
MAPK	mitogen-activated protein kinase
MBP	myelin basic protein
mCPBA	meta-chloroperbenzoic acid
MeCN	acetonitrile
MEF	murine embryonic fibroblast
MESNA	2-mercaptoethanesulfonic acid
MPPA	(4-carboxymethyl)thiophenol
N-terminus	amino-terminus
Nle	norleucine
NCL	native chemical ligation
NPE	2-nitrophenylethyl
OD	optical density
PBS	phosphate buffered saline
PBS	paxillin binding subdomain
PCS	protease cleavage sequence
PDE	phosphodiesterase
PNT domain	pointed domain
Pfa	phosphono-difluoro-methylenealanine
PKL	paxillin kinase linker
Pma	phosphonomethylenealanine
Pmp	phosphonomethylene phenylalanine
PNA	polyamide nucleic acid
PTP	protein tyrosine phosphatase
pSer	phosphoserine
pThr	phosphothreonine
pTyr	phosphotyrosine
PyBOP	benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate
RBD	Ras binding domain
SDS-PAGE	sodium dodecyl-sufate polyacrylamide gel electrophoresis
SARA	Smad anchor for receptor activation
SeM	selenomethionine
SH2	Src homology 2
SPPS	solid phase peptide synthesis
$t_{1/2}$	half life
TBS	tris buffered saline
TEV	tobacco etch virus
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIRF	total internal reflection fluorescence
TMS	trimethylsilane
TNBS	trinitrobenzene sulfonic acid
UV	ultraviolet

Chapter 1. Introduction

NCL: Semisynthesis of post-translationally modified proteins and biological probes

A version of this chapter has been submitted for publication in *Protein Engineering*.

1-1. Introduction to native chemical ligation

Methods for the modification of proteins to introduce novel chemical functionalities and for the preparation of homogenous samples of post-translationally-modified proteins are critical for probing protein structure/function relationships and protein/protein interactions. Native chemical ligation is a chemoselective reaction that joins synthetic or biologically-expressed protein fragments *via* a native amide bond, enabling the construction of modified proteins in multi-milligram quantities sufficient for biophysical and biochemical studies.[1] Native chemical ligation can be carried out with fully deprotected peptide and protein fragments in neutral aqueous media, enabling modified peptides to be incorporated into target peptides or proteins. Required for the reaction are an *N*-terminal fragment containing a *C*-terminal α-thioester, and a *C*-terminal fragment with an *N*-terminal cysteine residue. The general reaction is shown in Fig. 1.1a. The transformation begins with a reversible transthioesterification to associate the two fragments *via* a thioester bond involving the cysteine residue of the *C*-terminal fragment. An energetically favorable acyl rearrangement involving a 5-membered ring transition state then occurs to form a stable amide bond. Since the thioester-linked intermediates have never been observed, it is presumed that the rate-limiting step of NCL is transthioesterification. Typically, the reaction is run in the presence of thiol additives, which both suppress cysteine oxidation and catalyze the reaction by generating more reactive thioesters.[2] Importantly, NCL is compatible with the presence of fully-deprotected internal cysteine residues in either protein fragment (Fig. 1.1b). Thioester exchange is fully reversible, and it is only upon reaction with the terminal cysteine that the S → N acyl rearrangement can occur, resulting in the thermodynamically stable product and accounting for the chemoselective nature of the reaction.

NCL was first introduced in 1994 with the synthesis of a 72-residue polypeptide comprising entirely native peptide bonds, thus overcoming the practical limits of SPPS through the ligation of two fully deprotected peptide fragments.[1] While NCL was originally limited to ligations involving α-thioesters generated by chemical synthesis, the scope of the reaction was advantageously extended to proteins of any size through the application of 'expressed protein ligation' (EPL), in which α-thioester or *N*-terminal Cys-containing fragments that are generated *via* recombinant methods are used for the semisynthesis of protein domains and full-length proteins.[3-5] EPL was originally defined specifically as the extension of NCL that involves recombinant α-thioester peptides, but the term is now used more generally to designate NCL processes involving any recombinant fragments, either thioester or *N*-terminal Cys-containing.[6]

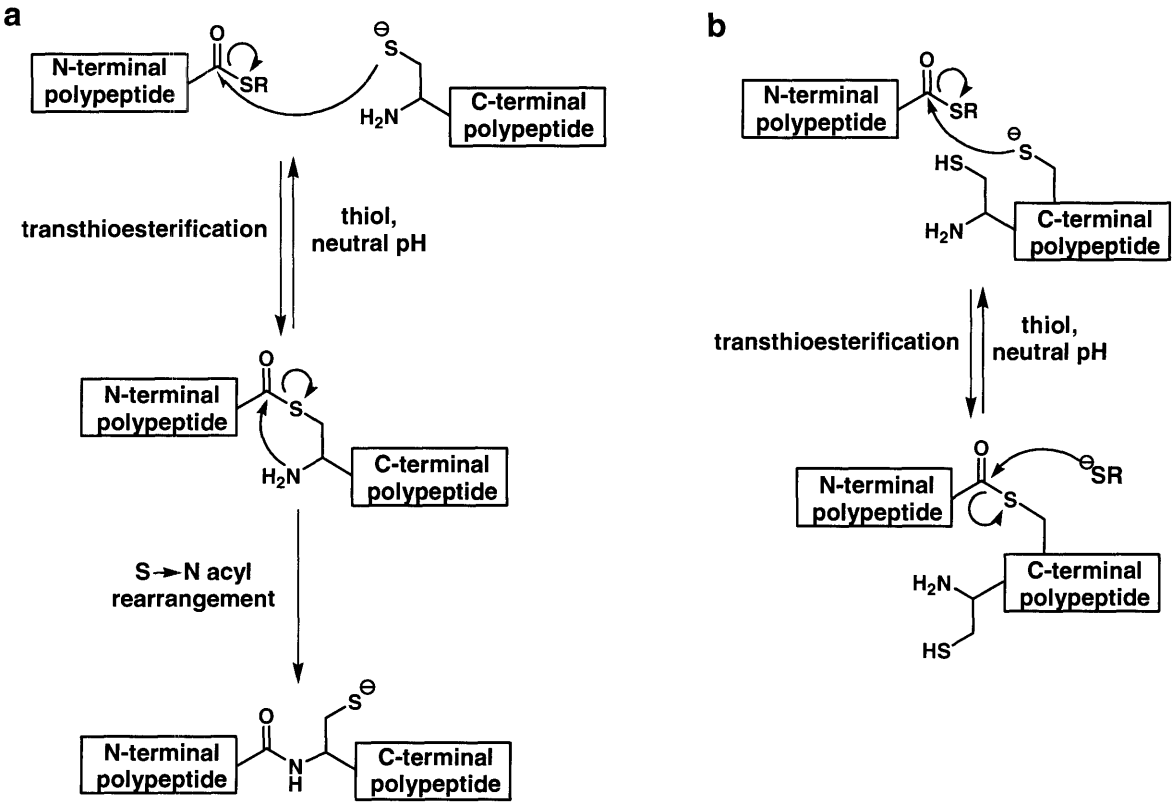


Figure 1.1. **a**) Native chemical ligation (NCL) of unprotected polypeptide fragments. **b**) Reversible transthioesterification with internal cysteine residues.

1.2 Engineering design considerations for NCL

An important consideration in the design of semisynthetic proteins is the designation of a ligation junction, Xaa-Cys. Since the practical limit of solid phase

peptide synthesis (SPPS) is between 40 and 60 residues, junction sites for ligations involving a synthetic peptide must fall within those distances from the *N*- or *C*-terminus of the protein target. In some cases naturally-occurring cysteine residues are present in the protein of interest at a suitable position for the ligation. However, in the absence of a suitably placed cysteine, one may be introduced in the place of a non-essential residue. Appropriate residues may be selected based on structural knowledge or previous experiments that demonstrate that the residues are not involved in protein function. The ease of site-directed mutagenesis allows for further confirmation that the cysteine mutation does not significantly alter protein structure or function. In general, ideal ligation sites are in segments of a protein that lack secondary structure, such as in terminal regions, loop regions or linker regions between two domains. If the reaction is carried out under non-denaturing conditions, and if the protein will not be subjected to refolding, it is particularly important that the folded state of the protein is not disturbed by the ligation site, regardless of whether a naturally-occurring or an introduced cysteine residue is used as the Xaa-Cys junction.

While standard NCL dictates that a cysteine residue be present at the *N*-terminus of the *C*-terminal fragment, it is also important to consider the effect of the *C*-terminal residue of the α-thioester fragment, -Xaa-SR, on the ligation. It has been reported that α-thioester peptides containing any of the 20 encoded amino acid residues at the terminal position are compatible with NCL, but the kinetics of ligation differ dramatically depending on the properties of that residue.[7] In general, the reaction proceeds most rapidly when the *C*-terminal position is occupied by the sterically-unhindered glycine residue, or by histidine or cysteine residues, which are hypothesized to increase the rate of thioester exchange by catalysis with the imidazole or thiol side-chain functionalities. Other terminal residues shown to result in rapid product formation are phenylalanine, methionine, tyrosine, alanine, and tryptophan. Ligation has been found to be prohibitively slow only with two of the β-branched amino acids (valine and isoleucine) and proline, although it may be possible to overcome these residue limitations using highly-activated thioesters.

In another study, ligations involving aspartic or glutamic acid residues in the *C*-terminal position of an α-thioester were shown to result in significant side product

formation.[8] In this case, the proximity of the carboxyl functionality in the amino acid side chains to the activated thioester can result in the formation of backbone isomers, in which a β– or γ– amide bond forms between the Cys-containing peptide and the aspartate or glutamate residues respectively (Fig. 1.2a). Side product formation can be avoided by orthogonal protection of the carboxyl side chains. However, when possible, selecting an Xaa-Cys ligation site that avoids sluggish ligations or side reactions is beneficial to ensure maximal product formation.

Another factor affecting the efficiency of NCL is the presence of thiol additives,[2] which potentially combat difficult ligation reactions. The alkylthioesters commonly used in NCL react slowly because of the poor leaving group properties of the corresponding alkyl thiols, requiring thiol-thioester exchange with the thiol additive to promote ligation with the Cys-containing fragment. (To minimize confusion, this exogenous thiol replacement will be referred to throughout the chapter as thiol-thioester exchange to distinguish it from the attack of a Cys residue in the first step of NCL, which will be referred to as transthioesterification.) Rapid and complete ligation is best facilitated by thiols that are both good nucleophiles, to promote the *in situ* formation of a more reactive thioester, and good leaving groups, to favor the transthioesterification (Fig. 1.2b). For NCL involving peptide fragments, a benzylmercaptan/thiophenol mixture is commonly employed in an aqueous/organic buffer (Fig. 1.2c). Thiophenol promotes rapid thiol-thioester exchange and serves as a good leaving group, but has poor aqueous solubility and therefore can only be used at very low concentrations for ligations involving proteins that cannot tolerate the addition of organic solvent. MESNA (2-mercaptoethanesulfonic acid) has excellent solubility in water and is a popular choice for ligations using recombinant protein fragments. However, a recent investigation of a number of thiol additives revealed that MESNA, an alkanethiol with pKa of 9.2, shows rapid thiol-thioester exchange, but has poor leaving group properties.[9] In the case of MESNA, the rate limiting step for NCL is transthioesterification, whereas with certain other thiols the thiol-thioester exchange is rate limiting. In general alkylthioesters, such as those generated by MESNA or benzyl mercaptan, were found to be less reactive than phenylthioesters. The investigators found (4-carboxymethyl)thiophenol (MPPA) to be a superior catalyst with peptide-based test ligations and recommend use of MPPA for

protein ligations as well, predicting that it may be more effective than MESNA for EPL. Since this report was published within one month of writing this chapter, there have been no other reports to date that apply MPPA, and it will be interesting to see if MESNA is eventually replaced as the predominantly used thiol for EPL by MPPA or another aryl thiol catalyst. Another option for rapid ligation is the use of preformed aryl thioesters,[1,10,11] which eliminate the thiol-thioester exchange step that is rate limiting with certain thiol additives.

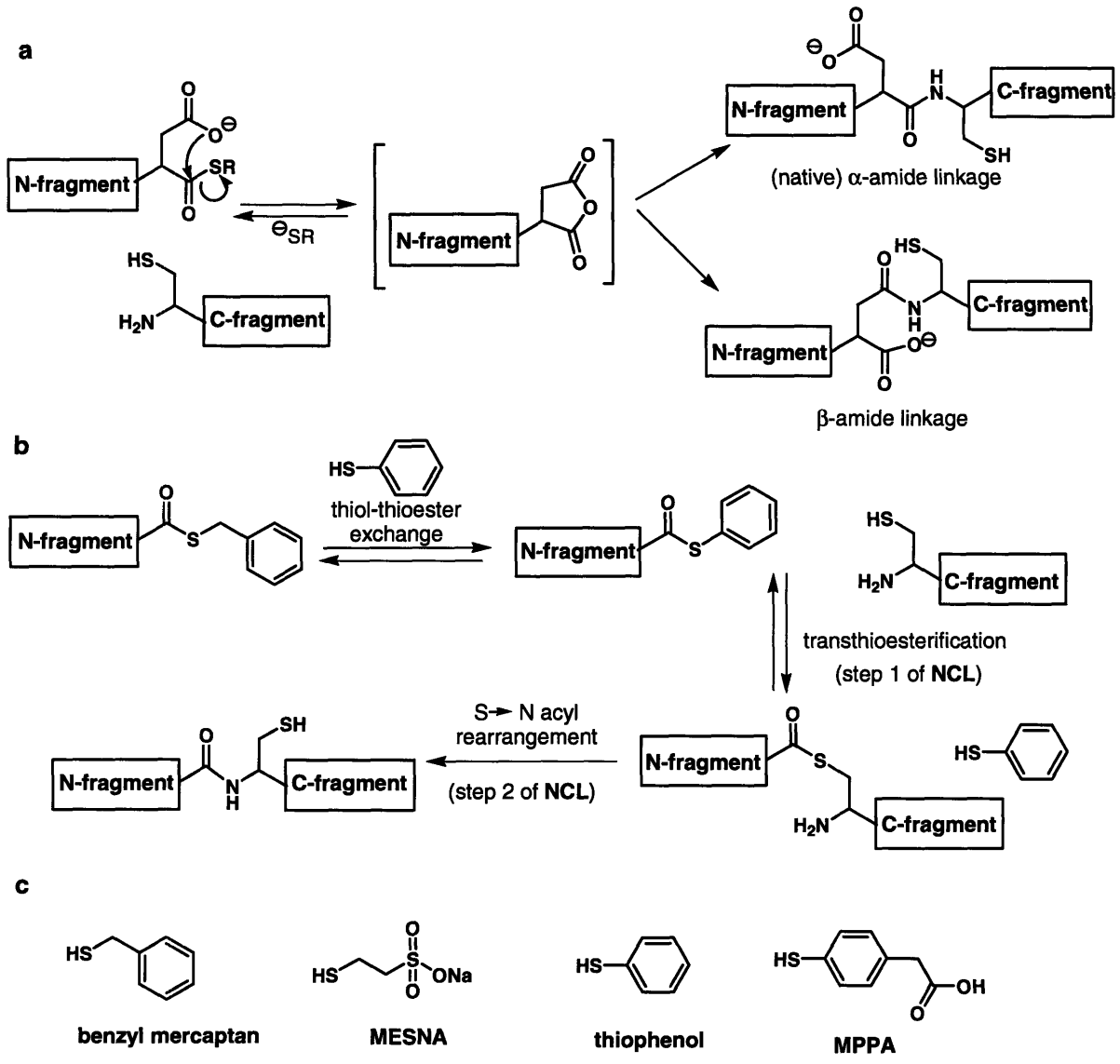


Figure 1.2. **a**) Generation of a β-amide-linked side product for NCL with an Asp-Cys ligation junction. **b**) Mechanism for NCL with thiol catalysis. An ideal thiol additive results in rapid thiol-thioeser exchange of the thioester fragment and provides a good leaving group for transthioesterification with the Cys-fragment. **c**) Thiols used for NCL.

1.3 Thioester generation

1 Synthetic thioester peptides

The synthesis of α-thioester peptides for NCL can be accomplished using a variety of SPPS-based methods. The modular nature of SPPS allows facile replacement of any residue of the target peptide thioester with non-encoded amino acids, as well as the site-specific incorporation of tags and probes. Initially most investigators employed acid-labile Boc-protection strategies for SPPS,[1,12] since the thioester linkage is not stable to the basic deprotection treatment with piperidine used in Fmoc-based SPPS. However, a number of approaches have been developed to circumvent the sulfur-carbonyl cleavage associated with Fmoc deprotection, allowing the generation of thioester peptides without the harsh cleavage conditions employed in Boc-based SPPS. This use of milder conditions is particularly important for the synthesis of glycopeptide and phosphopeptide thioesters, since glycoside and phosphoryl linkages are labile to the anhydrous HF required for peptide cleavage in Boc-based SPPS.

In one Fmoc-based approach, deblocking agents were developed as alternatives to piperidine to effectively remove the Fmoc protecting group without nucleophilic cleavage of the thioester linkage.[13,14] In an alternative strategy, a backbone amide linkage (BAL) is used to anchor the penultimate residue, which contains an orthogonally protected *C*-terminus, to the solid support.[15] The peptide can be elaborated using standard Fmoc-based SPPS protocols, and subsequently deprotected at the *C*-terminus prior to being coupled to a thioester amino acid and finally fully deprotected and cleaved from the resin (Fig. 1.3a). Additional BAL and side-chain linkage-based approaches have been reported,[16] including an example that masks the terminal glycine thioester as a trithioortho-ester and releases the deprotected thioester by acid treatment.[17]

A more commonly applied strategy utilizes an alkanesulfonamide "safety-catch" linker, in which a peptide is assembled on solid support linked via the *C*-terminus by a sulfonamide bond. This sulfonamide linkage is stable to the repeated piperidine deprotection treatments in Fmoc-based SPPS. Following peptide synthesis, the sulfonamide is activated by *N*-alkylation and subsequently cleaved from solid support by nucleophilic attack of a thiol additive. The thioester bond is stable in TFA, and peptide can therefore be deprotected with a standard TFA cleavage cocktail (Fig. 1.3b).[18,19] A

similar method employs an aryl hydrazide linker that can be activated following peptide synthesis by mild oxidation.[20]

Arguably the most straightforward Fmoc-based SPPS approach for generating thioesters involves peptide synthesis on highly acid-labile resin, such as commercially available 2-chlorotrityl or TGT resin, followed by cleavage of the fully-protected peptide from solid support with mild acid treatment. The acid releases the peptide as a *C*-terminal carboxylic acid, but does not affect the Fmoc-compatible side-chain protecting groups. The protected peptide is then derivatized to a *C*-terminal thioester *in situ* by treatment with the desired thiol and standard peptide-coupling activating agents.[21,22] Side-chain deprotection affords the corresponding free thioester (Fig. 1.3c). An evaluation of activating agents for this strategy has identified conditions, specifically phosphonium-salt based activating agents, that result in high yields and low levels of epimerization (<1.4%).[10]

While still in a preliminary stage of development, a final method of α-thioester formation is worth mentioning both for its creative approach and its similarity to the protein splicing mechanism that is exploited for the generation of recombinant protein α-thioesters (Sect. 1.3.2). This method involves the generation of a thioester-linked peptide by utilizing an N→S acyl shift facilitated by a protected thiol-containing auxiliary attached to the peptide backbone (Fig. 1.3d).[23] Following Fmoc-based SPPS of a peptide linked to resin by an amide bond, deprotection of the peptide side chains with a TFA cleavage cocktail concurrently deprotects the thiol moiety of the auxiliary, initiating an acyl shift that results in a thioester bond-linkage of the peptide to solid support. Subsequent thiolysis releases the corresponding peptide thioester.

A recent review of methods for thioester synthesis can provide more information on these and other approaches.[24]

2. Recombinant thioester fragments

While direct synthesis is suitable for peptide fragments under approximately 60 residues in length, the scope of NCL was initially limited by the lack of recombinant methods to generate longer peptide or protein fragments containing *C*-terminal α-thioesters. This limitation was overcome in 1998 with the introduction of EPL, which

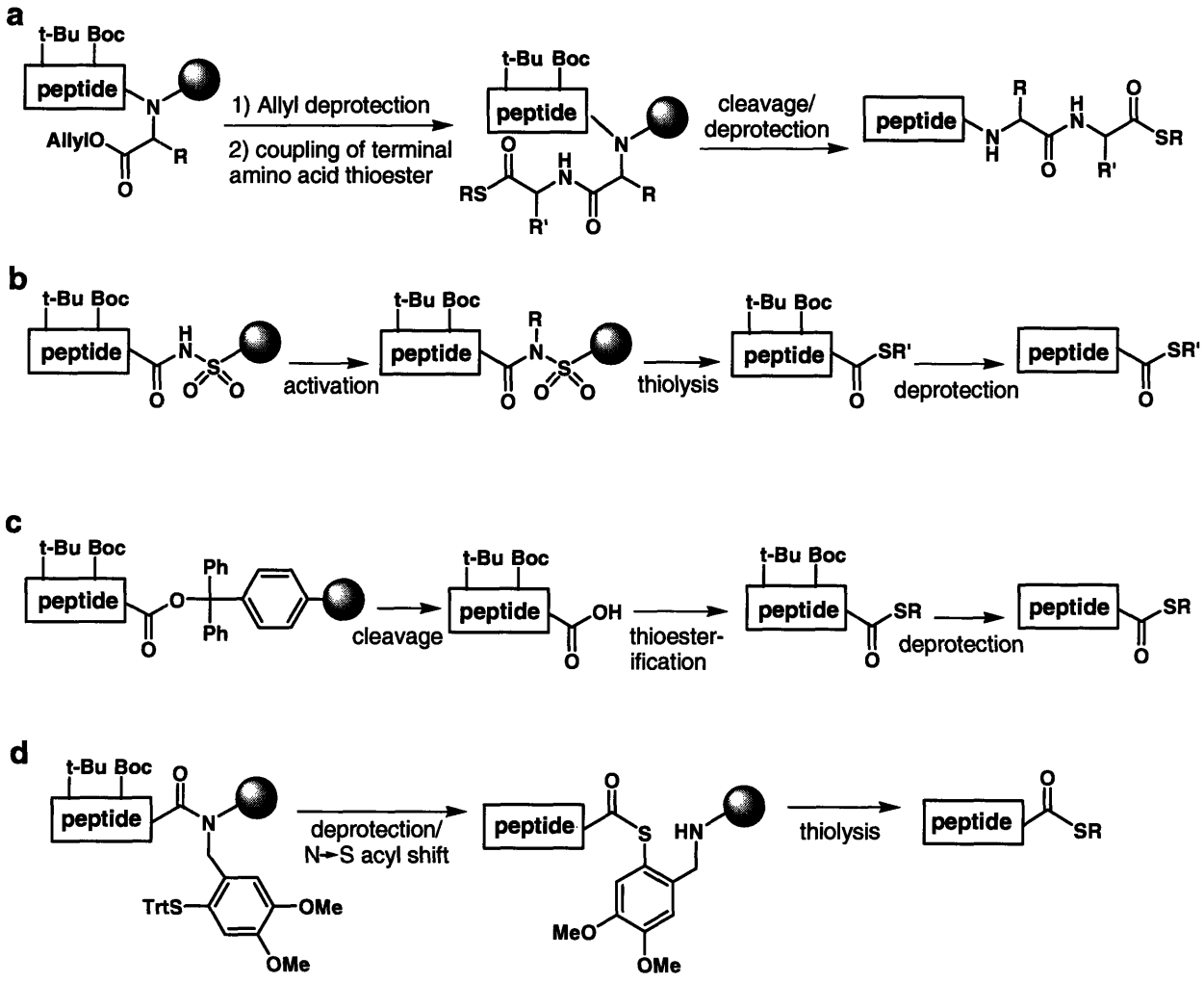


Figure 1.3. Fmoc SPPS-based methods for peptide α-thioester synthesis include the **a**) BAL strategy, **b**) "safety catch" linker strategy, **c**) in situ thioesterification of a protected peptide, and **d**) thiolysis following a thiol-auxiliary mediated N → S acyl shift.

applies a modified protein splicing mechanism to generate thioesters suitable for ligation.[3,4] To place the technology developed for recombinant protein thioesters in context, a brief explanation of protein splicing follows.

Protein splicing Protein splicing is a protein processing event that results in the extrusion of an internal protein segment, termed an intein, and the concomitant joining of two flanking regions, termed *N*- and *C*-exteins, through an amide bond.[25] Inteins, which are functionally analogous to self-splicing RNA introns, are protein segments that

catalyze the intramolecular protein rearrangement that mediates their excision (Fig. 1.4a). A number of conserved or key residues in inteins contribute to the splicing event through structural or electronic influences.[26] In general, inteins contain a cysteine or serine residue at the *N*-terminal position (termed the 1 position), a conserved asparagine residue at the *C*-terminal position, and a cysteine, serine, or threonine residue at the first position of their flanking *C*-extein (termed the +1 position). In the initial step of standard protein splicing, an N→S (or O) acyl shift occurs, involving the cysteine (or serine) at position 1, transferring the *N*-extein from the backbone to the side chain of residue 1 of the intein. While this step appears thermodynamically unfavorable, it is favored by the conformation of the intein, which is thought to distort the scissile amide bond into a higher energy conformation. In the next step, transesterification involving the cysteine (or serine or threonine) at position +1 joins the two exteins through a thioester bond. In the third step, the asparagine residue cyclizes, resulting in cleavage at the *C*-terminal splice junction to liberate the intein as a *C*-terminal succinimide. In the final step, which is the sole reaction that does not require catalysis by the intein, a spontaneous acyl rearrangement produces the spliced exteins with an amide linkage.[6]

Mutant (Asn to Ala) inteins to generate recombinant thioesters Protein splicing has been exploited with great success for the generation of recombinant thioesters. Inteins have been engineered with an asparagine to alanine mutation, which permits the initial step of the protein splicing mechanism involving an N → S acyl shift to produce a thioester linkage, but prevents subsequent succinimide formation.[27] Addition of exogenous thiols results in release of the *N*-extein as the corresponding *C*-terminal thioester. The isolation of recombinant thioesters from a resin-bound intein system has been commercialized by New England Biolabs as the IMPACT™ (intein-mediated purification with an affinity chitin binding tag) –system. In this system, a target gene, functioning as the *N*-extein, is cloned immediately *N*-terminal to a genetically modified (Asn to Ala) intein gene. A chitin binding domain (CBD) is cloned *C*-terminal to the intein, functioning as the *C*-extein and facilitating immobilization of the resulting protein construct on chitin beads. Thus, the expressed three-segment (target protein-intein-CBD)

construct can be isolated from all other cellular proteins by immobilization and washing, and the target protein thioester can be released by subsequent thiolysis (Fig. 1.4b).[28]

Expressed protein ligation can be performed with the eluted thioester or directly on the chitin beads.[3,4,29] For solid phase ligation, the thiol and Cys-containing peptide/protein fragment can be simultaneously incubated with the resin-bound protein fusion, enabling concurrent thiolysis and ligation.

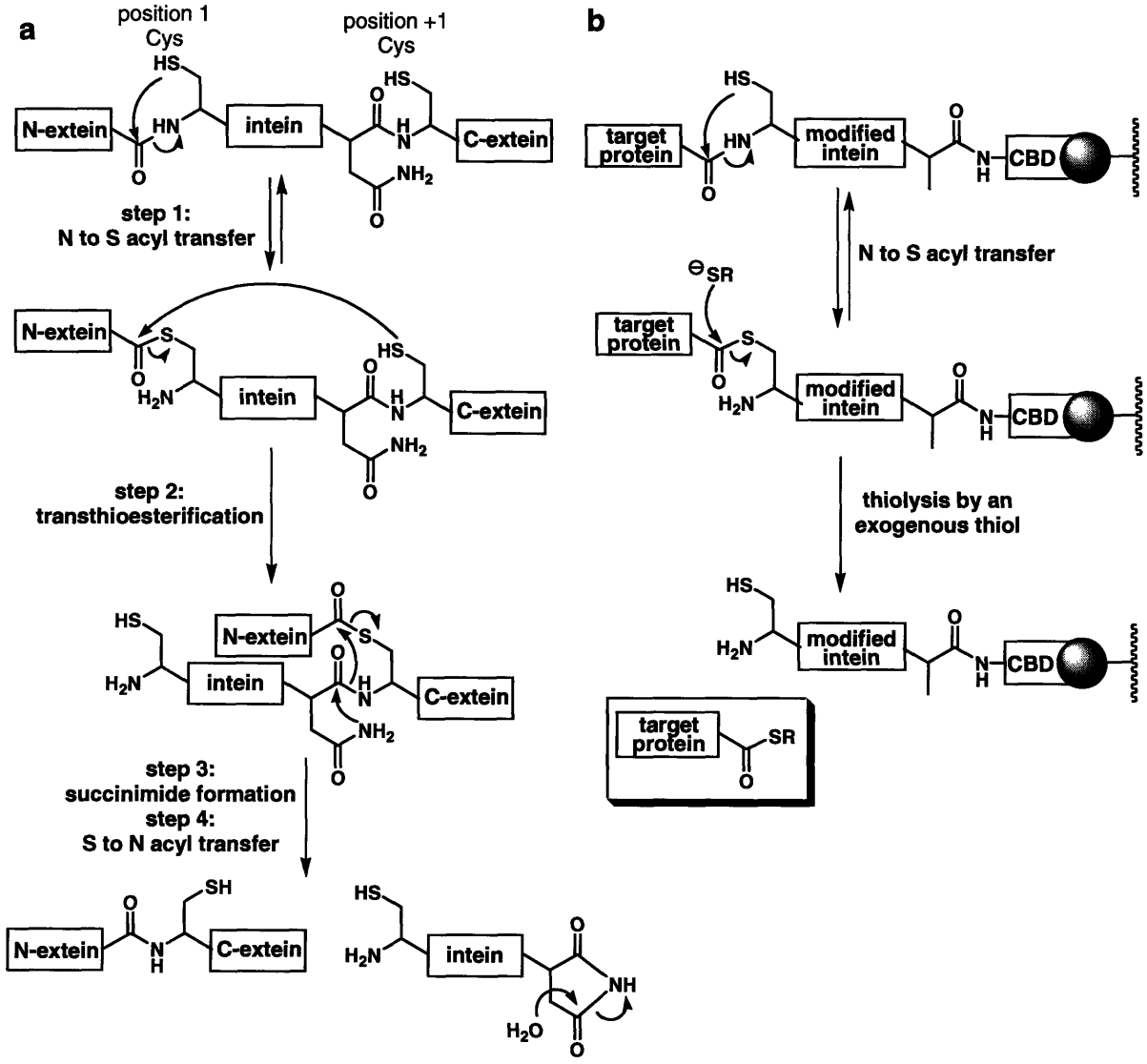


Figure 1.4. **a)** Mechanism of protein splicing. Splicing results in the joining of two exteins through a native peptide bond and the extrusion of the intein segment. **b)** Recombinant generation of an α-thioester using a modified intein system. The target protein is *N*-terminal to the intein, and a chitin binding domain (CBD) is *C*-terminal to the intein, immobilizing the construct on solid support. An Asn to Ala mutation prevents succinimide formation. Thiolysis results in release of the α-thioester target protein.

1-4. N-terminal Cys fragments

The synthesis of peptides containing an *N*-terminal cysteine residue is straightforward using SPPS and requires no additional manipulations. For longer fragments, there are several approaches for accessing biochemically-expressed proteins with an *N*-terminal cysteine. The most commonly applied methods involve the generation of a precursor protein designed with a cysteine residue immediately *C*-terminal to a protease cleavage site (Fig. 1.5a). The first example of affinity cleavage for NCL used factor Xa, a protease that cuts *C*-terminal to its Ile-Glu-Gly-Arg recognition sequence.[5] Other *C*-terminal cleaving proteases include enterokinase, ubiquitin C-terminal hydrolase, and furin. The single disadvantage with employing proteases is the possibility for undesired cleavage at secondary sites. For instance, factor Xa can cleave after Gly-Arg pairs or other basic residues in a target protein.

Recently tobacco etch virus (TEV) protease, a highly specific cysteine protease with a seven-residue recognition sequence, was applied for the generation of *N*-terminal Cys fragments.[30] TEV typically recognizes a Ser or Gly residue in the P1' site, but will also tolerate a Cys in that position. TEV demonstrates high sequence selectivity and overexpresses well on a large scale, making it an ideal protease for EPL applications.

Endogenous Met aminopeptidases have also been utilized to access *N*-terminal Cys proteins from the corresponding Met-Cys containing precursor proteins (Fig. 1.5b). The resulting Cys-polypeptides can be isolated from cell lysate using aldehyde-functionalized resin,[31] or reacted directly for *in vivo* NCL.[32] For exogenous cleavage of a Met-Cys junction, cyanogen bromide (CNBr) was successfully applied to access an *N*-terminal Cys in a recombinant glycoprotein that was insoluble in buffers compatible with the commonly employed proteases.[33]

An intein-based strategy can also be applied to generate *N*-terminal Cys proteins (Fig. 1.5c). Several commercially available expression vectors contain genetically modified inteins that lack the conserved cysteine (or serine) residue at the *N*-terminus (1 position) of the intein for this purpose. Cleavage of the intein by succinimide formation, induced by lowering the pH and increasing temperature, simultaneously releases the *N*-terminal Cys protein.[28] The limitation to this method is the possibility for spontaneous intein cleavage.

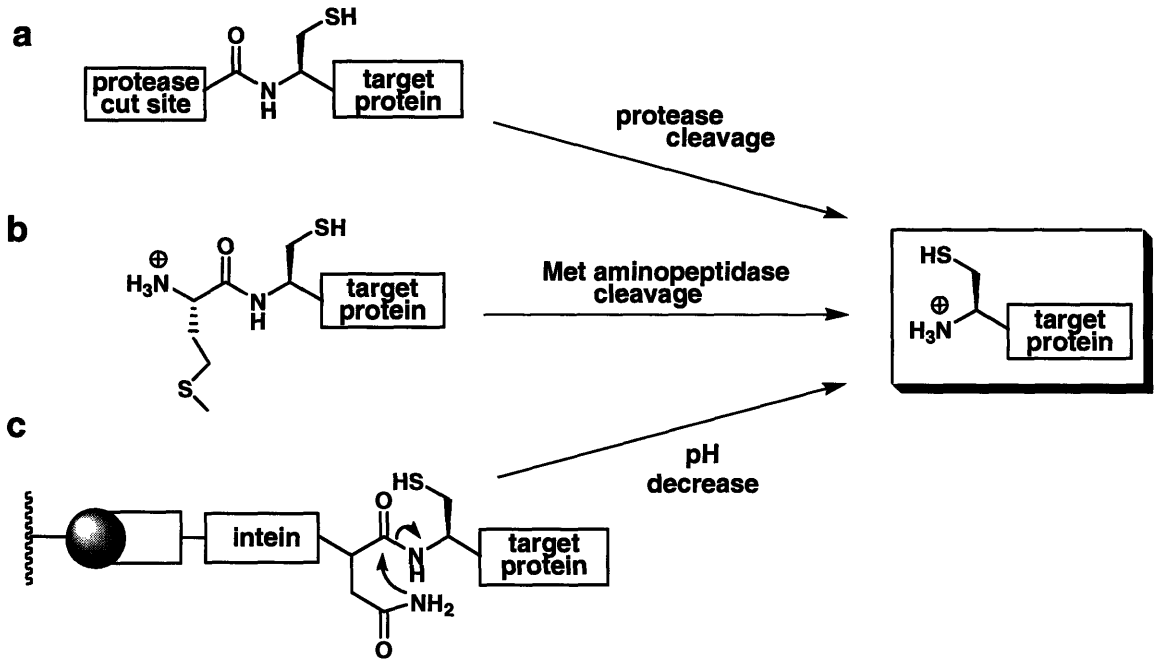


Figure 1.5. Recombinant methods for generating *N*-terminal Cys-containing protein fragments using **a**) exogenous protease, **b**) endogenous methionine aminopeptidase, and **c**) a genetically modified intein system.

1-5. Extensions of NCL

1. Sequential NCL

Although many NCL applications require the ligation of only two fragments, the technique is not limited to a single ligation step. Multiple modifications throughout a protein or modification in the middle of a large protein can be accomplished by sequential ligations (Fig. 1.6a).[34] In sequential NCL, an initial ligation is carried out between an *N*-terminal Cys-containing fragment and a second fragment containing a thioester moiety and a protected *N*-terminal Cys residue, masked to prevent the thioester fragment from undergoing intra- or intermolecular ligations. Strategies used to mask the cysteine residue include the incorporation of the factor Xa pro-sequence immediately *N*-terminal to the cysteine residue,[34] acetamidomethyl (Acm) protection of the side-chain thiol,[35] thiazolidine cysteine protection,[36] and photolabile protection of the thiol or amino group.[37] Following the initial ligation, the cysteine is deprotected and a second ligation is carried out. Theoretically there is no limit to the number of fragments that can be ligated

in this fashion. Segmental ligation has been performed on solid support, in a strategy analogous to SPPS,[38] as well as *in situ*, either in "one pot" reactions [37,39] or with HPLC or affinity-tag purification of the ligated intermediates.[39]

2. Accessing Xaa-Ala and Xaa-Gly ligation junctions

The only major limitation of NCL is the general requirement for a cysteine residue at the ligation site. For specific applications, the requisite *N*-terminal Cys residue can be replaced by another nucleophilic residue, such as selenocysteine.[40] Alternatively, NCL can be combined with post-ligation desulfurization with palladium or H_2/Raney nickel to convert the resultant cysteine to an alanine, and thereby access peptides or proteins with an Xaa-Ala ligation junction.[41] This method has been applied to several proteins that lack cysteine residues,[42,43] but cannot be used for proteins that contain a cysteine anywhere in the protein sequence, since all thiols will be reduced by the desulfurization step.

In a more generally applicable strategy, NCL can be performed with a removable Cys-mimic to ultimately produce an Xaa-Gly ligation junction (Fig. 1.6b).[44] In this approach the *C*-terminal ligation fragment is linked *via* the *N*-terminal amine to a thiol-containing auxiliary that facilitates thioester exchange with the α-thioester fragment. Following transthioesterification, an S → N acyl shift results, and subsequent cleavage of the *N*-linked auxiliary yields a native amide bond at the ligation site. Rapid ligation is best promoted by auxiliaries that proceed through a five-membered ring acyl rearrangement intermediate, and most of these utilize a 2-aryl mercaptoethyl group as the Cys-mimic. Depending on the aryl substituents, removal of the auxiliary is accomplished with strong acid treatment,[45] milder TFA treatment,[46] or photolysis.[47]

1.6 Applications

The strongest testament to the utility and generality of NCL is the impressive number of applications that have been reported for the study of complex biological

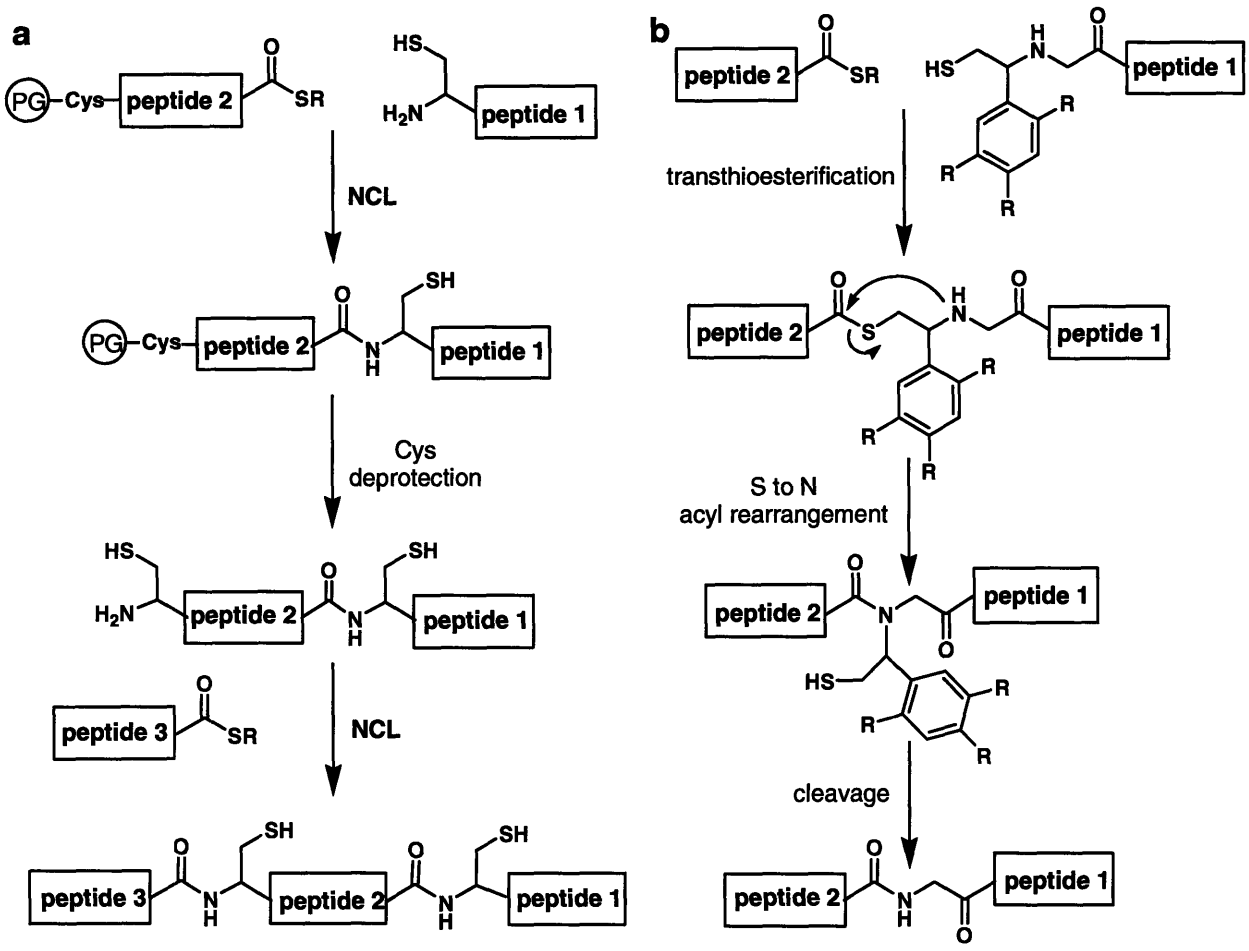


Figure 1.6. **a**) Sequential NCL. PG = Cys protecting group. **b**) NCL with a 2-aryl mercaptoethyl ligation auxiliary to produce an Xaa-Gly ligation junction.

systems using ligation methodology to access homogenous samples of native or modified proteins or protein analogs. Due to the vast number of applications reported in the twelve years since NCL was first introduced, only a fraction of these can be covered with adequate detail in this chapter. Since a number of excellent reviews are available,[6,26,48] particular focus is on reports from the past three years.

1. Post-translational modifications

Post- (and co-) translational modification of proteins increases the diversity of the proteome by more than an order of magnitude beyond that programmed by the genetic code.[49] Direct characterization of the impact of post-translational modifications,

including phosphorylation, glycosylation, lipidation, and acetylation, on the structure and function of proteins can be prohibitively complex due to the difficulty of obtaining homogenously modified protein in high yields. The nature of this difficulty stems from the lack of genetic encoding for these modifications, the heterogeneity of biological samples, and the inherent non-specificity in the enzymes that catalyze post-translational modifications. NCL and EPL allow the generation of modified proteins in milligram quantities by the reaction of a biologically-expressed non-modified fragment with a synthetic fragment containing the desired modification. Many post-translational modifications occur within the *N*- or *C*-terminus of proteins, making the corresponding modified proteins ideal targets for NCL. However, segmental ligation can be used to incorporate modifications within the central regions of a protein as well.

Phosphorylation Phosphorylation is the most prevalent post-translational modification, affecting an estimated one third of human proteins.[49] Protein kinases, enzymes which catalyze transfer of the γ-phosphoryl from ATP to a serine, threonine, or tyrosine side chain within a peptide or protein, can have tens or hundreds or substrates, and may modify multiple sites in a single protein, making the production of discretely phosphorylated proteins challenging. NCL has facilitated the preparation of homogeneous samples of phosphoproteins, enabling investigators to isolate the specific roles of phosphorylation on protein function *in vitro*.

In one example, synthetic variants of a Cys_2His_2 zinc finger protein were prepared by NCL in various phosphorylated forms and used to study the effects of linker phosphorylation on DNA binding.[50] Cys_2His_2 zinc finger proteins are a class of transcription factors comprising zinc-binding domains joined by highly conserved linker regions, which each include a single threonine or serine residue. These linker regions are known to be phosphorylated, but previous studies on the effects of phosphorylation were limited to using partially-purified protein isolated from cell lysates, prohibiting detailed quantitative evaluation. NCL enabled the preparation of pure phosphorylated variants of a representative 86-residue protein containing three zinc finger domains joined by two linker regions. The proteins were synthesized by sequential ligation of peptide fragments to access variants with either a single phosphothreonine residue in the *N*-terminal or *C*-

terminal linker, phosphothreonine residues in both linkers, or no linker phosphorylation (Fig. 1.7a). Direct comparison of the DNA-binding affinity of the three phosphorylated variants and the non-phosphorylated protein using a fluorescence anisotropy–based assay indicated that phosphorylation of either linker resulted in a ~ 40-fold decrease in DNA binding affinity, and that phosphorylation of both linkers resulted in a 130-fold loss. These quantitative measurements, made possible by pure preparations of phosphorylated variants, provide strong support for a model the authors set out to evaluate, in which coordinated regulation of zinc finger transcription factors is achieved by cellular phosphorylation of the linker regions.

In another study, singly- and dually-phosphorylated variants of a signaling protein, Smad2, were prepared using EPL, allowing investigators to deconvolute the impact of each phosphoryl group on protein oligomerization and on further Smad2 phosphorylation by an upstream kinase.[51] Phosphorylation of receptor-activated Smad (R-Smad) proteins, such as Smad2, occurs on the final two serine residues of a *C*-terminal phosphorylation sequence. Phosphorylation of this *C*-terminus results in the dissociation of the R-Smad from the membrane-bound receptor complex, and the formation of a new heteromeric complex between the R-Smad and a related co-Smad protein, which subsequently translocates to the nucleus and regulates gene expression. Phosphorylated R-Smads can also form homotrimers *in vitro*. With the serine residues under investigation located at the extreme *C*-terminus, Smad2 and the three possible phosphorylated variants were ideal candidates for EPL and were prepared by ligation of an expressed α-thioester corresponding to Smad2 (residues 241-462) with one of four synthetic pentapeptides containing a phosphoserine residue at either position 465 or 467, both 465 and 467, or at neither position (Fig. 1.7b). Biophysical studies on the oligomerization state of the variants, including analytical ultracentrifugation, revealed that stable Smad2 oligomer formation requires phosphorylation at both serine residues but that phosphorylation of Ser465 provides the driving force for oligomerization. In addition, phosphorylation of Smad2 at Ser467 proceeds more rapidly when the protein is already phosphorylated at Ser465, while the rate of phosphorylation of Smad2 at Ser465 does not significantly increase by prephosphorylation at Ser467. Since Smad2 is enzymatically phosphorylated on both sites by the same kinase, a semisynthetic strategy

was critical to enable access to pure samples of singly phosphorylated Smad2 for these studies. Interestingly, the receptor kinase that phosphorylates Smad2, TβRI, is itself activated by phosphorylation and a tetraphosphorylated variant of this kinase was among the first applications of NCL to produce phosphoproteins[52,53] and was utilized in the Smad2 phosphorylation assay.

Glycosylation Homogenous samples of glycosylated proteins are particularly challenging to access because cellular glycoproteins exist as complex mixtures of glycoforms, which are difficult to purify or even characterize.[54] Similar to phosphorylation, enzymatic glycosylation results in heterogeneous samples due to modification of multiple residues, and glycoprotein mixtures may be further complicated by the diversity of oligosaccharides and possible sugar linkages. Tremendous effort has been devoted to the chemical and chemoenzymatic synthesis of glycopeptides,[55] and these advances have been applied to the synthesis and semisynthesis of homogenous glycoproteins by NCL.[19,56-59] For the preparation of glycopeptide α-thioesters, Fmoc-based SPPS strategies are exclusively applied, since glycosidic linkages are not stable to the repetitive acid treatments required in Boc-based SPPS.

In an impressive synthetic undertaking, a 132–residue mucin-like glycoprotein, GlyCAM-1, was semisynthesized in three distinct glycoforms containing up to 13 glycans for future investigations on the effect of glycosylation on GlyCAM-1 structure and function.[35] GlyCAM1 consists of a central unglycosylated domain flanked by two mucin domains, characterized by dense groups of α-*O*-linked *N*-acetylgalactosamine (GalNAc) on serine and threonine residues. All glycans were introduced as the mono-saccharide GalNAc derivative, which, in principle, can be elaborated enzymatically to produce fully active glycoprotein. The semisynthetic glycoforms generated in this study include GlyCAM-1 variants glycosylated exclusively in either the *N*-terminal or *C*-terminal mucin domain and a variant glycosylated in both mucin domains. The authors applied three different NCL approaches to achieve all the glycoforms, including a sequential strategy using two ligations and involving a recombinant α-thioester with a masked *N*-terminal cysteine to generate the GlyCAM-1 variant glycosylated on both mucin domains (Fig. 1.7c).

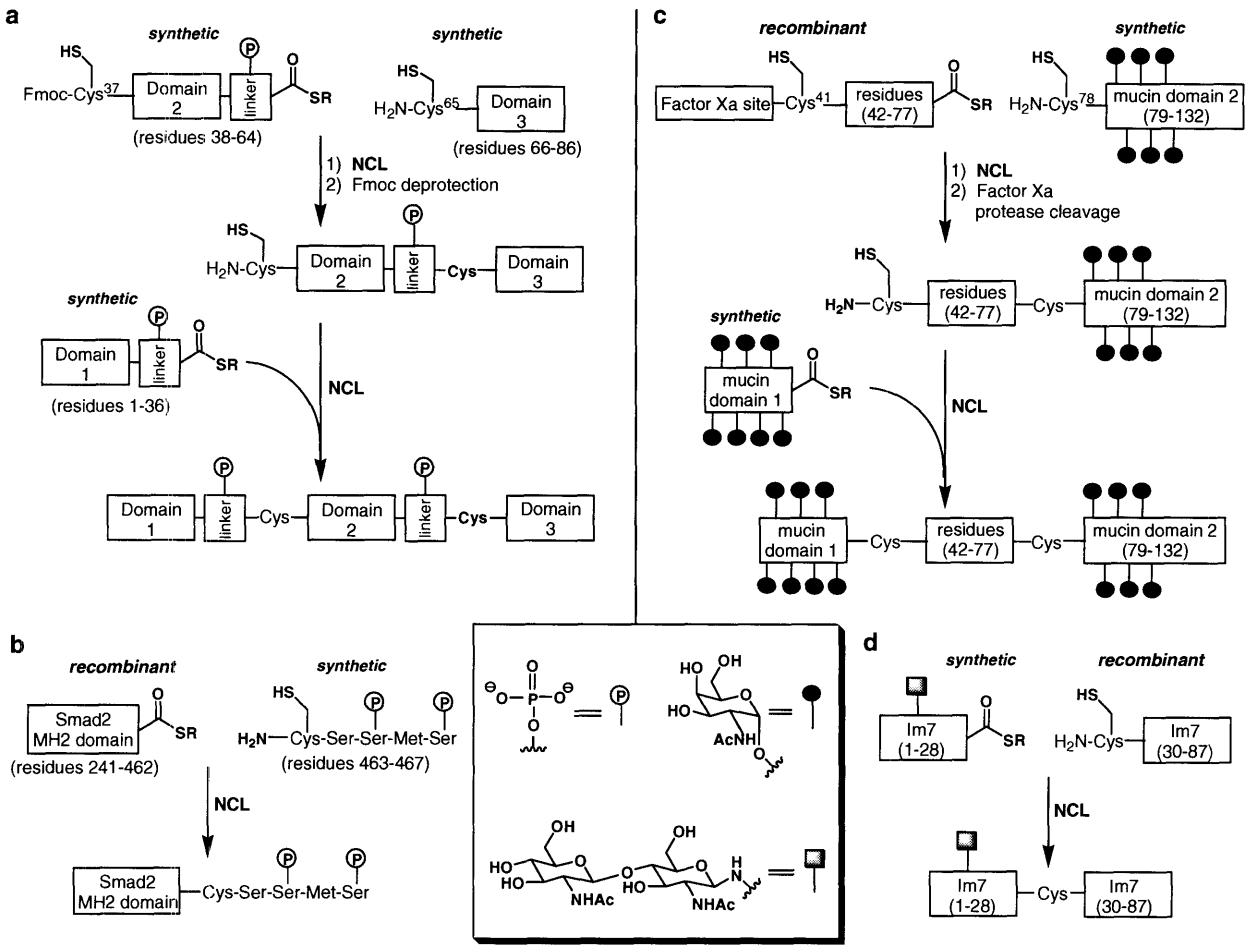


Figure 1.7. Semisynthesis of **a**) a dually-phosphorylated zinc finger protein, **b**) Smad2 phosphorylated on residues 465 and 467 of the *C*-terminal tail, **c**) GlyCAM-1 with 13 GalNAc modifications on two mucin-like domains, and **d**) an N-linked glycoprotein variant of Im7.

NCL has also been applied to the semisynthesis of an *N*-linked glycoprotein variant of a well-studied bacterial protein, Im7, to investigate the influence of glycosylation on protein folding.[60] *N*-linked glycosylation is a co-translational event that is thought to assist in the correct folding of expressed proteins. The four-helix Im7 protein, which is not naturally glycosylated, served as a tractable model for folding studies due to the significant quantity of data available on the kinetics and thermodynamics of its three-state folding mechanism. The glycoprotein analogue was prepared with an Asn-linked chitobiose building block at residue 13 on helix I by the

ligation of a glycosylated α-thioester to an expressed *C*-terminal fragment comprising Im7 residues 29-87 (Fig. 1.7d). Biophysical analysis of the glycosylated and non-glycosylated semisynthetic Im7 variants revealed that glycosylation at position 13 had minimal effect on protein folding. Investigations of other Im7 glycovariants are currently underway to probe the effect of glycosylation at other sites.

Lipidation The post-translational lipidation of proteins is involved in regulating function by targeting modified proteins to specific membranes. The covalent addition of lipid anchors can be divided into four main classes: *N*-terminal myristoylation, *C*-terminal addition of a glycosyl phosphatidylinositol (GPI), acetylation, and prenylation.[49] As with other post-translationally modified proteins, lipoproteins are challenging to access by genetic or enzymatic methods and have become exciting targets for NCL.

Prenylation involves the addition of a farnesyl or geranylgeranyl group to one or two cysteines at the *C*-terminus of a protein. The *C*-terminal location of the modified cysteine residues renders prenylated proteins well suited for semisynthesis by EPL. A successful application has been the semisynthesis of fluorescently-labeled mono- and diprenylated variants of Rab7, a Rab guanosine triphosphatase (GTPase) of the Ras-GTPase superfamily.[61] An expressed α-thioester corresponding to Rab7 (residues 1-201) was reacted with fluorescently-labeled and prenylated hexapeptides to generate Rab7 variants with geranylgeranyl modifications on either Cys205 or Cys 207, or on both Cys205 and Cys207 of the *C*-terminus (Fig. 1.8a). Semisynthesis involving the hydrophobic lipopeptides required several modifications to standard EPL protocol, including the addition of specific detergents and an organic extraction of unreacted peptide, which binds to the protein non-covalently, from precipitated protein. In addition, a Rab chaperone protein was necessary to stabilize the denatured Rab7 variants during refolding. A novel fluorescence-based prenylation assay, utilizing the environment-sensitive properties of the dansyl fluorophore, was developed to probe the mechanism of diprenylation. The results of the prenylation assay support a proposed random sequential mechanism of prenylation. The straightforward assay, facilitated by the creative application of a fluorescent tag, was possible because the researchers were able to generate homogenous monoprenylated protein.

As with glycosylation, a primary challenge in the EPL of prenylated proteins is the synthesis of modified peptides. Comparison of a number of solution-phase and solid-phase methods for synthesizing prenylated peptides revealed the strength of a hydrazide linker-based solid-phase approach that was used to incorporate several prenylated and fluorescently labeled peptides onto the α-thioester fragment of Rab7.[62] The extensive semisynthetic work with Rab7 was recently extended to other Ras-type GTPases: those in the Ras subfamily.[63] Prenylated Ras proteins pose additional semisynthetic challenges because they are not known to interact with a chaperone, the use of which was essential for the stabilization and purification of Rab7. This challenge was successfully addressed by the use of polybasic prenylated peptides, which eliminated non-specific peptide/protein aggregation and enabled ligation and purification in non-denaturing conditions.

Progress has also been made in the preparation of proteins with GPI anchors. GPI proteins are modified *via* a *C*-terminal amide linkage with a lipo-pentasaccharide anchor. Using EPL, lipidated analogs of green fluorescent protein (GFP) were created with a simplified GPI anchor, a phospholipid without glycans, to demonstrate a flexible strategy for generating proteins lipidated at the *C*-terminus.[64] The lipidated GFP variants were shown to incorporate stably into supported membranes, and quantification of their lateral fluidity was achieved by fluorescence imaging. This strategy can therefore be applied to the semisynthesis of naturally lipidated proteins and their study in lipid bilayers.

Acetylation Reversible acetylation involves modification of proteins, notably histones and transcription factors, on the ε-amino group of lysine residues. In core histones, which comprise the octomeric protein core of nucleosomes, post-translational modification of the *N*-terminus alters histone-DNA interactions and is implicated in regulating gene transcription.[49] NCL has been employed to generate pure samples of modified histones, acetylated or methylated on the *N*-terminal tail.[42,65] In a recent example, a homogenous monoacetylated variant of histone H4 was prepared by NCL to characterize the structural and functional effects of acetylation of Lys16.[66] Toward this end, a synthesized peptide thioester acetylated at Lys16 and corresponding to residues 1-22 of histone H4 was ligated to a recombinant *C*-terminal fragment (residues 23 to 102)

(Fig. 1.8b). The H4 variant was incorporated into nucleosomal arrays and found to inhibit the formation of higher order chromatin structures and prevent the functional interaction of histones with a specific chromatin-associated protein. This characterization of a selectively acetylated variant complements previous peptide competition studies and provides direct evidence not accessible with truncated or randomly hyperacetylated histone derivatives.

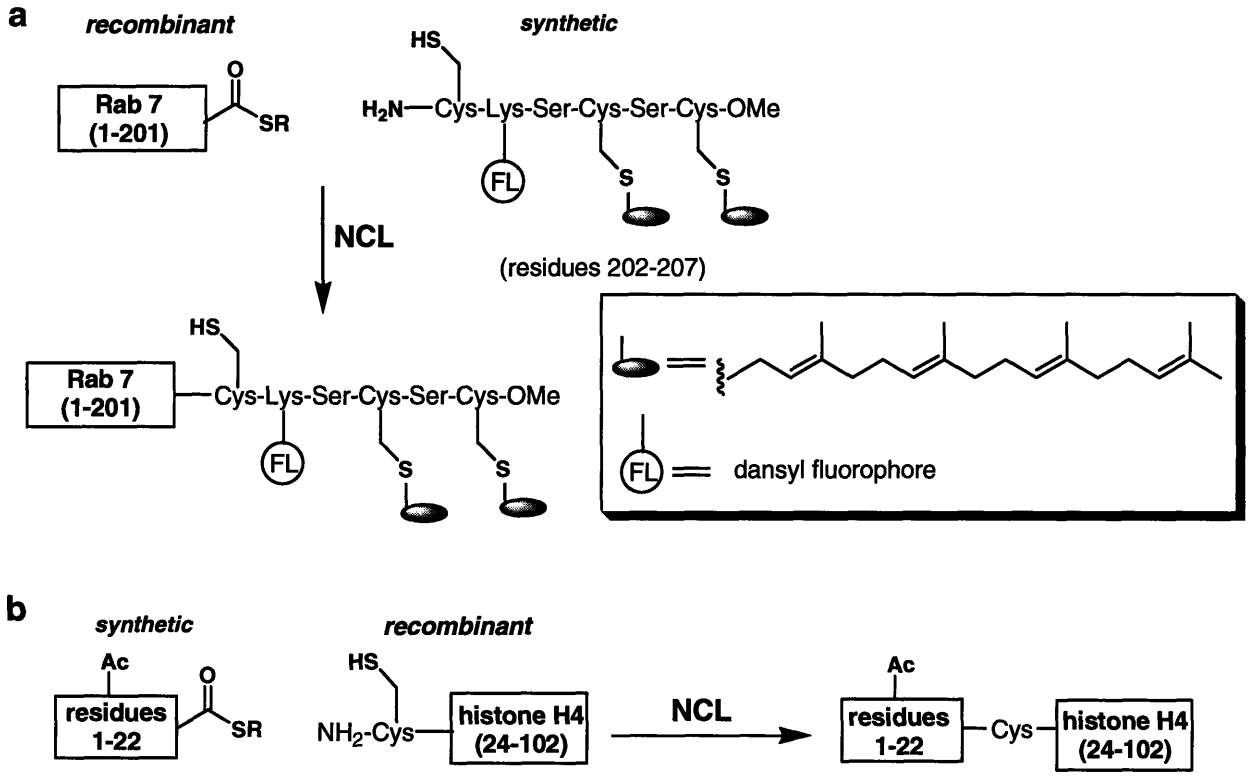


Figure 1.8. Semisynthesis of **a**) fluorescently-labeled and diprenylated Rab7, and **b**) monoacetylated histone H4.

2. Fluorescent probes

Genetically-encoded fluorophores are widely used for imaging proteins in live cells. Green fluorescent protein (GFP) and the ever increasing number of GFP variants with improved properties and a spectrum of excitation and emission wavelengths are invaluable probes for imaging protein localization and examining intramolecular and intermolecular protein interactions. One powerful application of fluorophore-labeled

proteins is fluorescence resonance energy transfer (FRET). FRET results when two fluorophores are within close proximity and the emission spectrum of the "donor fluorophore" overlaps with the excitation spectrum of the "acceptor fluorophore." The proximity and spectral overlap enable a transfer of energy and a corresponding increase in the intensity of the acceptor fluorophore emission, allowing for quantification of the distance between the two fluorophores. The major drawback of using genetically-encoded fluorophores for FRET and other fluorescence-based imaging is the significant size of the fluorophores (27 kDa for GFP) appended onto the protein of interest, potentially altering native interactions and localization. Organic fluorophores are significantly smaller (< 1 kDa) and often possess superior photophysical properties, such as higher extinction coefficients and greater resistance to photobleaching, but can be challenging to incorporate into proteins in a chemoselective manner.

NCL has been used to chemoselectively install a donor-accepter pair of organic fluorophores into proteins to study both intramolecular conformational changes[38] and intermolecular interactions[67] using FRET. In an example of the latter, variants of serotonin *N*-acetyltransferase (AANAT), a circadian rhythm enzyme, were constructed *via* EPL with a fluorescein or rhodamine-containing peptide at the *C*-terminus.[67] Since standard methods for determining oligomerization state, such as size-exclusion chromatography and dynamic light scattering, proved inconclusive with AANAT, FRET was employed to probe for homo-oligomerization (Fig. 1.9a). Fluorescence studies showed a significant increase in FRET upon incubation of the donor (fluorescein) and acceptor (rhodamine)-containing AANAT variants, indicating a preference for AANAT to oligomerize.

Along with FRET, there have been numerous applications reported for fluorescent probes incorporated into proteins by NCL, some of which are discussed in the context of other sections in this chapter. Examples include the *C*-terminal labeling of proteins for anisotropy-based binding studies,[68] the domain-specific replacement of tryptophan residues with a red-shifted and environment-sensitive tryptophan analog to probe domain function in the context of a full-length protein,[69] and the site-specific incorporation of an environment-sensitive fluorophore into an effecter domain to monitor protein-domain interactions.[70] Fluorescent labeling of proteins has also been accomplished *in vivo*.[71] A

cell permeable fluorescent thioester was introduced into cells expressing a target protein with an *N*-terminal cysteine, generated by intein-mediated splicing, resulting in NCL to generate a protein variant labeled at the *C*-terminus (Fig. 1.9b).

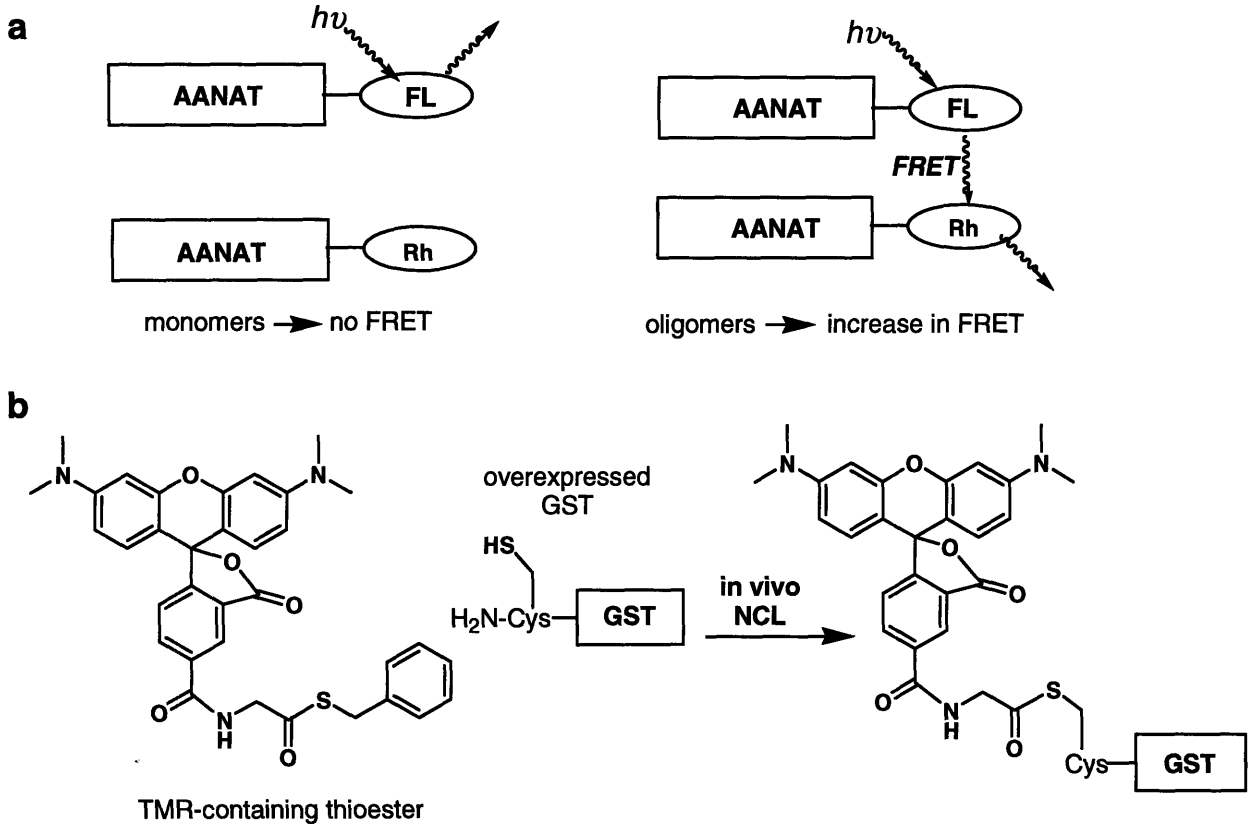


Figure 1.9. **a**) AANAT labeled with fluorescein (FL) and rhodamine (Rh). Expected FRET outcome for AANAT as a monomer (no FRET) and oligomer (FRET). **b**) In vivo labeling of GST (containing an *N*-terminal Cys residue) with the membrane permeable TMR thioester.

3. Unnatural amino acids

The power of NCL is even more pronounced in its application to the semisynthesis of protein domains or full-length proteins containing unnatural amino acids. Amino acid analogs chemoselectively introduced by NCL can be used to probe specific aspects of amino acid or protein function. This approach can involve introducing residues that differ from the natural amino acid at a position of interest in a single aspect, such as side-chain geometry, steric effects, or electronic effects. In one example, variants

of Src, a substrate of the kinase Crk, were prepared, in which a tyrosine residue known to be phosphorylated was replaced with one of five tyrosine analogs in order to dissect the contribution of individual factors affecting tyrosine phosphorylation (Fig. 1.10a).[72] Evaluation of the semisynthetic substrates in a radioactivity-based kinase assay led the authors to conclude that the phenolic hydroxyl of tyrosine is not involved in ground-state Src/Crk interactions and that stabilization of tyrosine conformers increases the efficiency of phosphorylation. In another study, EPL was used to generate analogs of the blue copper protein azurin, an electron transfer protein, to study the role of the copper-coordinating methionine residue on reduction potential (Fig. 1.10b).[73] The methionine residue was replaced with either a norleucine (Nle) or a selenomethionine (SeM) using NCL; subsequent spectroscopic and biochemical studies of the semisynthetic variants revealed hydrophobicity to be the most significant factor affecting reduction potential.

Introduction of unnatural amino acids has also been used to confer proteins with new properties or functions. For example, a variant of the enzyme ribonucease A (RNase A) was engineered with a β-peptide module in place of two natural amino acids, generating an enzyme with increased conformational stability without compromising catalytic activity (Fig. 1.10c).[74] NCL has similarly been used to create proteins with more potent binding properties. Src, a protein substrate of Csk, was prepared by EPL with an ATPγS conjugate in place of the tyrosine at the site of phosphorylation, creating a Csk kinase inhibitor (Fig. 1.10d).[75] That report demonstrates a strategy that could potentially be used to identify unknown kinases through pull-down assays, or to generate tightly bound kinase/substrate pairs for X-ray crystallography. In another example in which unnatural amino acid introduction leads to modulation of protein binding, a DNA-binding zinc finger protein was tuned to recognize a specific DNA sequence by incorporating citrulline (Cit), an amino acid which combines the side-chain length of arginine with an alternative functional group (Fig. 1.10e).[76] While tandem zinc finger domains provide a flexible framework for creating discriminating DNA-binding proteins, the domains almost exclusively require a guanosine at the 5'-end of the DNA for specific binding. Replacement of the DNA-interacting arginine residue with citrulline, which provides both a hydrogen-bond donor and acceptor, created a variant that binds preferentially to

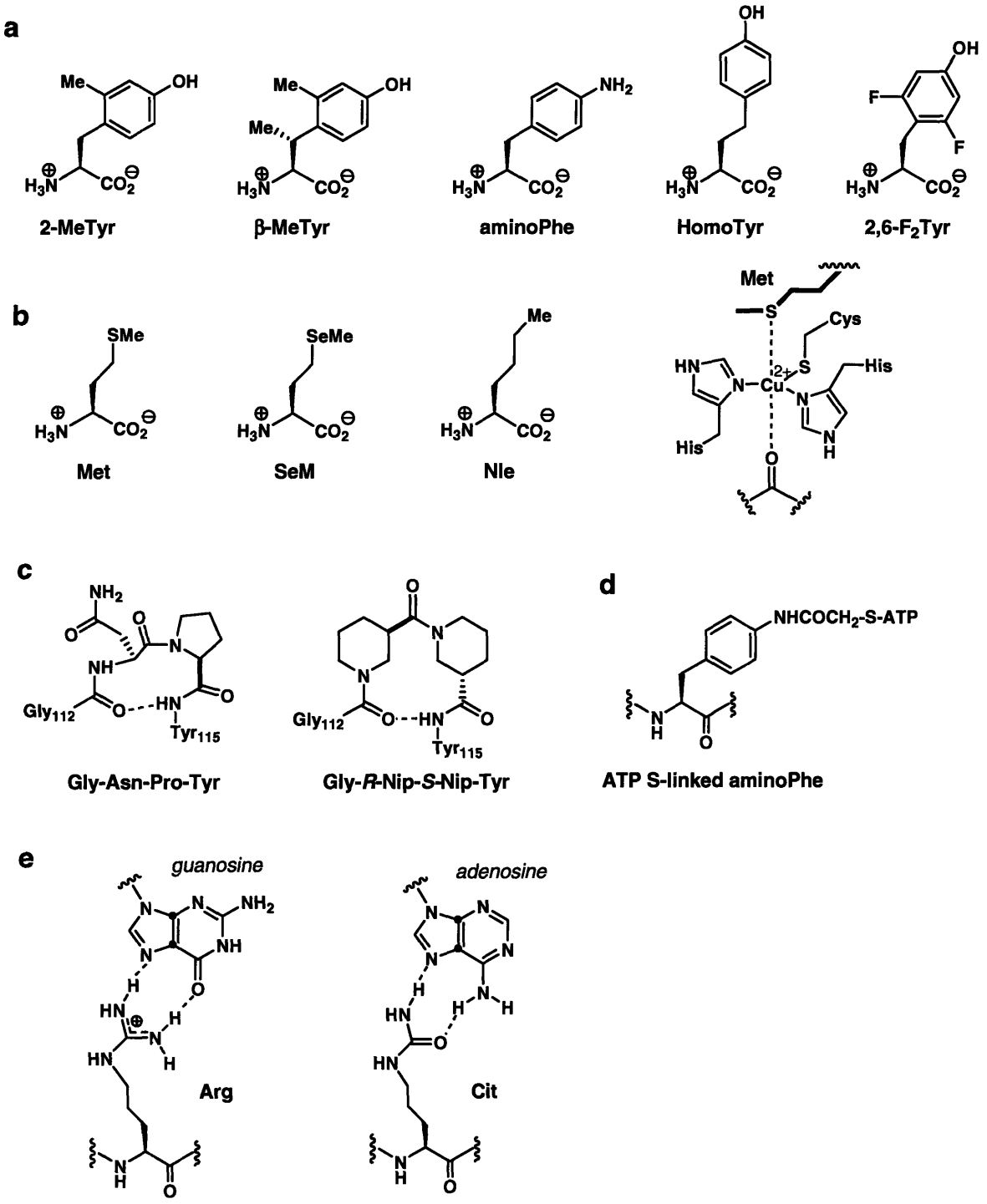


Figure 1.10. **a**) Unnatural tyrosine analogs incorporated into Src in place of a tyrosine residue known to be phosphorylated. **b**) Unnatural methionine analogs used to replace the active site methionine in azurin variants **c**) Naturally occurring residues (Gly-Asn-Pro-Tyr) forming a Type VI reverse turn in RNaseA and the replacement β-peptide-containing moiety (Gly-*R*-nipecotic acid-*S*-nipecoic acid). **d**) ATPγS-linked aminoPhe residue incorporated into Src to replace the tyrosine residue phosphorylated by the kinase Csk. **e**) arginine-guanosine interaction and the hypothesized Cit-adenosine interaction.

adenosine at the 5'-end of the DNA, a function not possible with zinc finger domain variants restricted to the 20 encoded amino acids.

The reports described above represent the use of unnatural amino acids to modulate or probe protein function but are by no means comprehensive. Several particularly creative applications that utilize NCL for the incorporation of unnatural amino acids are covered in detail below.

Phosphatase-resistant phosphoamino acid analogs Characterizing the role of protein phosphorylation *in vivo* is complicated by the presence of protein phosphatases, which are promiscuous enzymes that catalyze the removal of phosphoryl groups from phosphoproteins. Phosphatases can cause unanticipated and undetected hydrolysis of the phosphoprotein being studied. The phosphatases themselves can be modulated by phosphorylation, creating phosphoproteins challenging to study both *in vivo* and *in vitro*, since phosphatases have the propensity to autodephosphorylate. The introduction by EPL of nonhydrolyzable phosphotyrosine analogs into full-length protein has been used to great effect in the study of protein tyrosine phosphatase SHP-2 (Fig. 1.11a).[77,78] Two phosphatase-resistant phosphoTyr analogs were incorporated into the C-terminal tail (at either residue 542 or 580, or both 542 and 580) of SHP-2: phosphonomethylene phenylalanine (Pmp), which is commercially available as the *N*-Boc derivative, and difluoromethylene phosphonate (F_2Pmp), which is a superior mimic of phosphotyrosine but requires synthesis of the amino acid for introduction into peptides. Biochemical investigations of the SHP-2 variants revealed that phosphonates at Tyr542 and Tyr580 bind to *N*-terminal and *C*-terminal SH-2 domains respectively within SHP-2, disrupting basal inhibition and thereby increasing phosphatase activity. Near additive effects were observed for the doubly-phosphonylated mutant. Similar incorporation of Pmp into a related phosphatase, SHP-1, provided equally illuminating mechanistic insight into the role of tyrosine phosphorylation on phosphatase regulation.[79] More recently, single and double phosphonate substitutions were incorporated by EPL into the low molecular weight protein tyrosine phosphatase (LMW-PTP) and shown to inhibit dephosphorylation of phosphopeptide substrates, representing the first example of negative regulation of a tyrosine phosphatase by phosphorylation.[80] Also noteworthy, the effect of

phosphorylation on cellular stability and localization was investigated following the microinjection of the phosphonate variants into cells, an experiment that would not be possible with native phosphoproteins, which are susceptible to cellular phosphatases.

Phosphonate isosteres of phosphothreonine and phosphoserine, phosphonomethylenealanine (Pma) and phosphono-difluoro-methylenealanine (Pfa), have also been introduced into a full-length protein by NCL (Fig. 1.11b).[81,82] Unlike phosphotyrosine, which has no suitable surrogate within the encodable amino acids, glutamic and aspartic acid are commonly used as stable phosphoThr or phosphoSer mimics. However, these residues are not always successful mimics, as was reported with replacement of phosphoThr and phosphoSer residues in the protein AANAT, a regulatory enzyme involved in the production and secretion of melatonin. In contrast, the replacement of an *N*-terminal Thr31 or a *C*-terminal Ser205 with Pma by NCL enabled direct analysis of the role of phosphorylation on AANAT stability. Microinjection of the phosphonate mutants into cells provided direct support of increased AANAT stability by intermolecular binding facilitated by phosphorylation of Thr31 and Ser205.

Caged phosphorylated amino acids The caging, or photolabile protection, of biomolecules enables spatial and temporal control over the release of an activated species. The application of caged proteins holds great potential for elucidating protein function and signaling pathways but traditionally has been limited by the difficulty of preparing proteins containing masked functionalities. EPL was successfully applied to the semisynthesis of a caged variant of the signaling protein Smad2, which included two phosphoserine residues protected by the 2-nitrophenylethyl (NPE) photolabile protecting group (Fig. 1.11c).[83] Previous investigations by the same laboratory used homogenous phosphoSmad2 variants to explore the effect of phosphorylation on Smad2 oligomerization (Sect. 1.6.1).[51] The authors wished to expand on that work by using a dually-caged phosphoserine variant as a probe to study Smad2 pathway dynamics with spatial and temporal resolution in cellular experiments. To generate the probe, a peptide corresponding to the final five residues (463-467) of Smad2 and containing two caged phosphoserine residues was synthesized on the solid phase and ligated to a biologically-expressed α-thioester of the Smad2-MH2 domain (residues 241-462). Prior to

phosphorylation, Smad2 interacts with a membrane-bound complex, which includes 1:1 binding with a protein called SARA (**S**mad **a**nchor for **r**eceptor **a**ctivation). Upon phosphorylation, Smad2 releases from the SARA complex and can form homotrimers or heterocomplexes with a co-Smad, resulting in an active complex that localizes to the nucleus for gene expression regulation. The caged phosphoSmad2 variant was shown to mimic a nonphosphorylated Smad MH2 domain, forming a 1:1 complex with SARA. Upon irradiation, the release of the corresponding phosphodomain was detected, resulting in protein homotrimerization. In addition, a nuclear import assay was used to characterize the localization of the caged Smad2 variant before and after irradiation. As expected, prior to irradiation, caged Smad2-MH2 was excluded from the nucleus, similar to nonphosphorylated Smad2, while irradiation led to nuclear accumulation consistent with the release of phosphoSmad2. The caged phosphoSmad variant will be used in future experiments to provide quantitative information on Smad2 nuclear import and export in live cells.

Photocross-linkers Another recent application of EPL has been the semisynthesis of protein variants incorporating a photoactivatable crosslinker, introduced as benzoyl phenylalanine, to study a central binding interaction in the phototransduction cascade involved in vision.[84] The inactive form of cGMP phosphodiesterase (PDE) is known to be activated by nanomolar binding of the GTP-bound form of the α-subunit of transducin ($G\alpha_t$-GTP) with the PDE inhibitory γ-subunit (PDEγ). To further characterize this interaction, a series of seven full-length PDEγ photoprobes were constructed using EPL to install the benzophenone amino acid in place of selected *C*-terminal hydrophobic residues (Fig. 1.11d). The semisynthetic domains also included a biotin affinity tag at the *C*-terminus for analysis. Photocrosslinking experiments were performed between each of the PDEγ photoprobes and an activated form of $G\alpha_t$ ($G\alpha_t$-GTPγS). The investigators observed crosslinking with several of the photoprobes and were able to narrow down the region of photoinsertion into $G\alpha_t$, thereby revealing several previously unknown binding interactions.

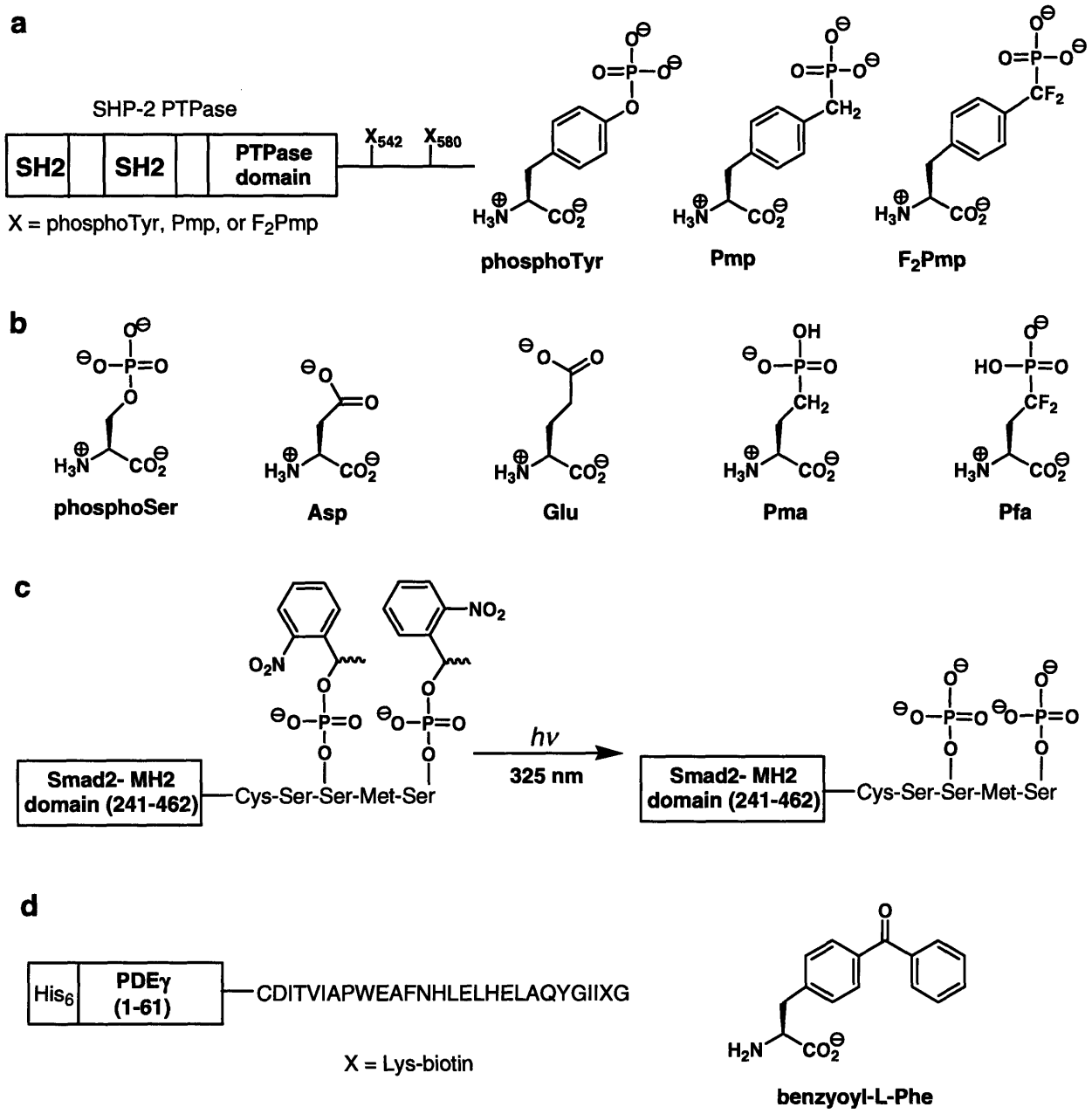


Figure 1.11. Applications of unnatural amino acids introduced into proteins by NCL. **a**) Replacement of Tyr residues 542 and 580 in the protein phosphatase (PTPase) SHP-2 with phosphatase-resistant Tyr mimics, Pmp and F_2Pmp. **b**) PhosphoSer and the genetically encoded phosphoSer/Thr mimics Asp and Glu, and the non-hydrolyzable phoshoSer/Thr mimics Pma and Pfa. **c**) Smad2-MH2 domain containing *C*-terminal caged phosphoSer residues. Irradiation with UV light releases the corresponding dually-phosphorylated Smad2 domain. **d**) PDEγ variants with the photocross-linking amino acid, benzoyl-L-Phe, incorporated in place of one of 7 hydrophobic residues highlighted in bold.

4. Probes for magnetic resonance: isotopic labeling and spin labeles

Structural analysis of proteins by nuclear magnetic resonance spectroscopy (NMR) is facilitated by the incorporation of carbon, nitrogen, and hydrogen isotopes, which may be uniformly introduced into biologically expressed proteins. At high molecular weights, despite isotopic labeling, NMR analysis becomes complicated due to decreased resolution, caused by longer correlation times, and highly complex spectra. EPL has been used to selectively introduce isotopic labels into protein domains within a full-length protein. This strategy has enabled NMR analysis of discrete protein domains or fragments within the context of large proteins, rendering a single section of the protein spectroscopically visible without disrupting domain-domain interactions. Specifically, EPL has been used to introduce a labeled domain at the terminus of a protein[85] and, using two sequential ligation reactions, to demonstrate the possibility of introducing an internal isotopically-labeled domain flanked by two non-labeled domains.[86]

In a recent example, the α subunit of a G protein (Gα) was semisynthesized with ^{13}C-labeled residues within the *C*-terminal tail to enable NMR characterization of conformational changes upon Gα subunit activation.[87] To construct the labeled full-length subunit, a synthetic nonapeptide (Gα residues 346-354) containing three ^{13}C-labeled amino acids was ligated to a recombinant thioester fragment corresponding to Gα residues 1-345. NMR characterization of the Gα subunit in its inactive (GDP-bound) form and in an active form (mimicked by the addition of AlF_4^-, a GTP γ-phosphate analog) provided evidence for an increase in the conformational order of the C-terminal tail upon Gα subunit activation.

EPL has also been employed for the incorporation of a paramagnetic amino acid into a full-length protein for electron paramagnetic resonance (EPR) spectroscopy.[88] A spin-labeled peptide containing a nitroxide-labeled lysine was ligated to a recombinant thioester of the Ras-binding domain (RBD) of *c*-Raf1. The ligation reaction, carried out in the presence of a thiol catalyst, resulted in reduction of the spin label, which was reoxidized prior to EPR analysis of the labeled RBD and a labeled RBD/Ras complex.

5. Immobilization by NCL and use in microarrays

The chemoselective nature of the NCL reaction has been exploited for immobilizing proteins on solid support in strategies aimed at the development of protein microarrays. Protein microarrays have the potential to facilitate high-throughput analysis of protein interactions, including protein/protein, protein/small molecule, and protein/antibody binding. A major challenge in the construction of protein microarrays is the uniform orientation of proteins on the glass support. In several examples, NCL has been used to install a reactive group at the terminus of proteins for subsequent immobilization. For example, researchers have used EPL to *C*-terminally label proteins with a biotin moiety by the ligation of the expressed protein thioesters with a biotin-labeled Cys residue.[89] The biotinylated proteins were subsequently immobilized on avidin-labeled glass slides, taking advantage of the stable biotin-avidin interaction. The immobilization of an oxidoreductase enzyme, AKR1A1, for future microarray applications was also accomplished using a similar biotin-labeling strategy.[90] In another immobilization strategy, researchers used EPL in the semisynthesis of protein-nucleic acid conjugates, in which recombinant thioester-containing proteins were ligated to peptides containing a *C*-terminal polyamide nucleic acid (PNA).[91] The proteins were then incorporated into small microarrays by DNA-directed immobilization (DDI) based on interaction of DNA on a glass slide to the complementary protein-conjugated PNA.

Proteins have also been attached to a solid surface through the stable amide bond formed directly by NCL (Fig. 1.12). Expressed protein α-thioesters have been shown to selectively immobilize onto a Cys-functionalized glass slide.[92] The reverse approach has also been accomplished, in which proteins containing *N*-terminal cysteines were immobilized onto thioester-containing slides.[93] In that case, the Cys-containing proteins were expressed using an intein-mediated strategy, and both the subsequently purified proteins and the crude cell lysates were used effectively to immobilize the *N*-terminal Cys-containing proteins onto the thioester-functionalized plates.

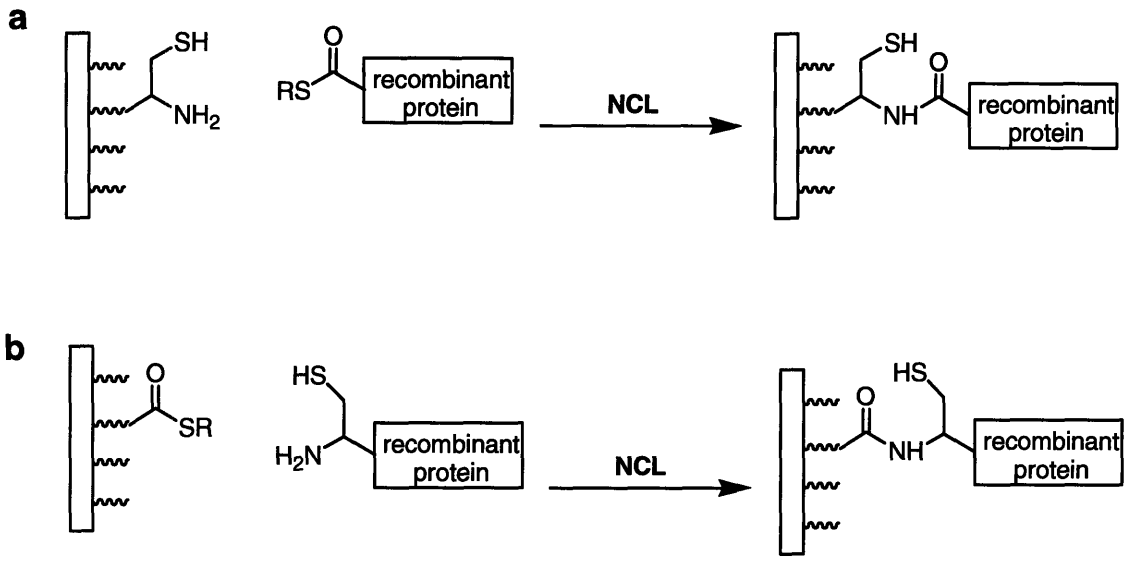


Figure 1.12. Immobilization of proteins onto microarrays by **a**) reaction of a thioester-containing protein with a Cys-functionalized slide and **b**) reaction of an *N*-terminal Cys-containing protein with a thioester-functionalized slide.

Conclusion

Advances in recombinant methods for the generation of α-thioester and *N*-Cys-containing proteins, and in synthetic approaches for peptide α-thioester synthesis, particularly methods compatible with peptides including posttranslational modifications and non-canonical amino acids, have facilitated the semisynthesis of proteins and protein domains containing unnatural amino acids and biological probes and tags in NCL applications. These advances have enabled investigators to design and construct elaborate protein biomolecules that can be used to address challenging problems not routinely accessible by traditional methods.

References

1. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science (Washington, D. C.)* **266**, 776-9 (1994).

2. Dawson, P. E., Churchill, M., Ghadiri, M. R. & Kent, S. B. H. Modulation of reactivity in native chemical ligation through the use of thiol additives. *J. Am. Chem. Soc.* **119**, 4325-4329 (1997).
3. Severinov, K. & Muir, T. W. Expressed protein ligation, a novel method for studying protein-protein interaction in transcription. *J. Biol. Chem.* **273**, 16205-16209 (1998).
4. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6705-6710 (1998).
5. Erlanson, D. A., Chytil, M. & Verdine, G. L. The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex. *Chem. Biol.* **3**, 981-991 (1996).
6. Muir, T. W. Semisynthesis of proteins by expressed protein ligation. *Annual Review of Biochemistry* **72**, 249-289 (2003).
7. Hackeng, T. M., Griffin, J. H. & Dawson, P. E. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10068-10073 (1999).
8. Villain, M., Gaertner, H. & Botti, P. Native chemical ligation with aspartic and glutamic acids as C-terminal residues: Scope and limitations. *Eur. J. Org. Chem.*, 3267-3272 (2003).
9. Johnson, E. C. B. & Kent, S. B. H. Insights into the mechanism and catalysis of the native chemical ligation reaction. *J. Am. Chem. Soc.* **128**, 6640-6646 (2006).
10. von Eggelkraut-Gottanka, R., Klose, A., Beck-Sickinger, A. G. & Beyermann, M. Peptide αthioester formation using standard Fmoc-chemistry. *Tetrahedron Lett.* **44**, 3551-3554 (2003).
11. Bang, D., Pentelute, B. L., Gates, Z. P. & Kent, S. B. Direct on-resin synthesis of peptide-αthiophenylesters for use in native chemical ligation. *Org. Lett.* **8**, 1049-1052 (2006).
12. Canne, L. E., Walker, S. M. & Kent, S. B. H. A general method for the synthesis of thioester resin linkers for use in the solid phase synthesis of peptide-α-thioacids. *Tetrahedron Lett.* **36**, 1217-20 (1995).

13. Li, X., Kawakami, T. & Aimoto, S. Direct preparation of peptide thioesters using an Fmoc solid-phase method. *Tetrahedron Lett.* **39**, 8669-8672 (1998).
14. Bu, X., Xie, G., Law, C. W. & Guo, Z. An improved deblocking agent for direct Fmoc solid-phase synthesis of peptide thioesters. *Tetrahedron Lett.* **43**, 2419-2422 (2002).
15. Alsina, J., Yokum, T. S., Albericio, F. & Barany, G. Backbone amide linker (BAL) strategy for Nα-9-fluorenylmethoxycarbonyl (Fmoc) solid-phase synthesis of unprotected peptide p-nitroanilides and thioesters. *J. Org. Chem.* **64**, 8761-8769 (1999).
16. Gross, C. M., Lelievre, D., Woodward, C. K. & Barany, G. Preparation of protected peptidyl thioester intermediates for native chemical ligation by Nα-9-fluorenylmethoxycarbonyl (Fmoc) chemistry: considerations of side-chain and backbone anchoring strategies, and compatible protection for N-terminal cysteine. *J. Pept. Res.* **65**, 395-410 (2005).
17. Brask, J., Albericio, F. & Jensen, K. J. Fmoc solid-phase synthesis of peptide thioesters by masking as trithioortho esters. *Org. Lett.* **5**, 2951-2953 (2003).
18. Ingenito, R., Bianchi, E., Fattori, D. & Pessi, A. Solid phase synthesis of peptide C-terminal thioesters by Fmoc/t-Bu chemistry. *J. Am. Chem. Soc.* **121**, 11369-11374 (1999).
19. Shin, Y. et al. Fmoc-Based Synthesis of Peptide-αThioesters: Application to the Total Chemical Synthesis of a Glycoprotein by Native Chemical Ligation. *J. Am. Chem. Soc.* **121**, 11684-11689 (1999).
20. Camarero Julio, A., Hackel Benjamin, J., de Yoreo James, J. & Mitchell Alexander, R. Fmoc-based synthesis of peptide alpha-thioesters using an aryl hydrazine support. *J. Org. Chem.* **69**, 4145-51 (2004).
21. Futaki, S., Sogawa, K., Maruyama, J., Asahara, T. & Niwa, M. Preparation of peptide thioesters using Fmoc-solid-phase peptide synthesis and its application to the construction of a template-assembled synthetic protein (TASP). *Tetrahedron Lett.* **38**, 6237-6240 (1997).

22. Mezo, A. R., Cheng, R. P. & Imperiali, B. Oligomerization of uniquely folded mini-protein motifs: development of a homotrimeric ββα peptide. *J. Am. Chem. Soc.* **123**, 3885-3891 (2001).

23. Kawakami, T., Sumida, M., Nakamura, K. i., Vorherr, T. & Aimoto, S. Peptide thioester preparation based on an N-S acyl shift reaction mediated by a thiol ligation auxiliary. *Tetrahedron Lett.* **46**, 8805-8807 (2005).

24. Camarero, J. A. & Mitchell, A. R. Synthesis of proteins by native chemical ligation using Fmoc-based chemistry. *Protein Peptide Lett.* **12**, 723-728 (2005).

25. Paulus, H. Protein splicing and related forms of protein autoprocessing. *Annual Review of Biochemistry* **69**, 447-496 (2000).

26. David, R., Richter Michael, P. O. & Beck-Sickinger Annette, G. Expressed protein ligation. Method and applications. *Eur. J. Biochem.* **271**, 663-77 (2004).

27. Chong, S. et al. Protein splicing involving the Saccharomyces cerevisiae VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. *J. Biol. Chem.* **271**, 22159-22168 (1996).

28. Chong, S. & Xu, M.-Q. in *Nucleic Acids and Molecular Biology* (ed. Gross, H. J.) 273-292 (Springer, Berlin Heidelberg New York, 2005).

29. Evans, T. C., Jr., Benner, J. & Xu, M.-Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci.* **7**, 2256-2264 (1998).

30. Tolbert, T. J. & Wong, C.-H. New methods for proteomic research: preparation of proteins with N-terminal cysteines for labeling and conjugation. *Angew. Chem., Int. Ed. Engl.* **41**, 2171-2174 (2002).

31. Villain, M., Vizzavona, J. & Rose, K. Covalent capture: a new tool for the purification of synthetic and recombinant polypeptides. *Chem. Biol.* **8**, 673-679 (2001).

32. Camarero, J. A., Fushman, D., Cowburn, D. & Muir, T. W. Peptide chemical ligation inside living cells: in vivo generation of a circular protein domain. *Bioorganic & Medicinal Chemistry* **9**, 2479-84 (2001).

33. Macmillan, D. & Arham, L. Cyanogen bromide cleavage generates fragments suitable for expressed protein and glycoprotein ligation. *J. Am. Chem. Soc.* **126**, 9530-9531 (2004).

34. Cotton, G. J., Ayers, B., Xu, R. & Muir, T. W. Insertion of a synthetic peptide into a recombinant protein framework: a protein biosensor. *J. Am. Chem. Soc.* **121**, 1100-1101 (1999).

35. Macmillan, D. & Bertozzi, C. R. Modular assembly of glycoproteins: Towards the synthesis of GlyCAM-1 by using expressed protein ligation. *Angew. Chem., Int. Ed. Engl.* **43**, 1355-1359 (2004).

36. Bang, D. & Kent, S. B. H. Protein synthesis: A one-pot total synthesis of crambin. *Angew. Chem., Int. Ed. Engl.* **43**, 2534-2538 (2004).

37. Ueda, S., Fujita, M., Tamamura, H., Fujii, N. & Otaka, A. Photolabile protection for one-pot sequential native chemical ligation. *ChemBioChem* **6**, 1983-1986 (2005).

38. Cotton, G. J. & Muir, T. W. Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. *Chem. Biol.* **7**, 253-261 (2000).

39. Bang, D. & Kent, S. B. H. His6 tag-assisted chemical protein synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5014-5019 (2005).

40. Gieselman, M. D., Xie, L. & van der Donk, W. A. Synthesis of a selenocysteine-containing peptide by native chemical ligation. *Org. Lett.* **3**, 1331-1334 (2001).

41. Yan, L. Z. & Dawson, P. E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **123**, 526-533 (2001).

42. He, S. et al. Facile synthesis of site-specifically acetylated and methylated histone proteins: reagents for evaluation of the histone code hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12033-8 (2003).

43. Bang, D., Makhatadze, G. I., Tereshko, V., Kossiakoff, A. A. & Kent, S. B. Total chemical synthesis and X-ray crystal structure of a protein diastereomer: [D-gln 35]ubiquitin. *Angew. Chem., Int. Ed. Engl.* **44**, 3852-3856 (2005).

44. Canne, L. E., Bark, S. J. & Kent, S. B. H. Extending the applicability of native chemical ligation. *J. Am. Chem. Soc.* **118**, 5891-5896 (1996).

45. Botti, P., Carrasco, M. R. & Kent, S. B. H. Native chemical ligation using removable Nα-(1-phenyl-2-mercaptoethyl) auxiliaries. *Tetrahedron Lett.* **42**, 1831-1833 (2001).

46. Macmillan, D. & Anderson David, W. Rapid synthesis of acyl transfer auxiliaries for cysteine-free native glycopeptide ligation. *Org. Lett.* **6**, 4659-62 (2004).

47. Marinzi, C., Offer, J., Longhi, R. & Dawson Philip, E. An o-nitrobenzyl scaffold for peptide ligation: synthesis and applications. *Bioorganic & Medicinal Chemistry* **12**, 2749-57 (2004).

48. Schwarzer, D. & Cole Philip, A. Protein semisynthesis and expressed protein ligation: chasing a protein's tail. *Curr. Opin. Chem. Biol.* **9**, 561-9 (2005).

49. Walsh, C. T. *Posttranslational modification of proteins: expanding nature's inventory* (Roberts and Company, Englewood, 2006).

50. Jantz, D. & Berg Jeremy, M. Reduction in DNA-binding affinity of Cys2His2 zinc finger proteins by linker phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7589-93 (2004).

51. Ottesen, J. J., Huse, M., Sekedat, M. D. & Muir, T. W. Semisynthesis of phosphovariants of Smad2 reveals substrate preference of activated TβRI kinase. *Biochemistry* **43**, 5698-5706 (2004).

52. Flavell Robert, R. et al. Efficient semisynthesis of a tetraphosphorylated analogue of the Type I TGFbeta receptor. *Org. Lett.* **4**, 165-8 (2002).

53. Huse, M., Holford, M. N., Kuriyan, J. & Muir, T. W. Semisynthesis of hyperphosphorylated type I TGFβ receptor: addressing the mechanism of kinase activation. *J. Am. Chem. Soc.* **122**, 8337-8338 (2000).

54. Macmillan, D. & Bertozzi, C. R. New directions in glycoprotein engineering. *Tetrahedron* **56**, 9515-9525 (2000).

55. Grogan, M. J., Pratt, M. R., Marcaurelle, L. A. & Bertozzi, C. R. Homogeneous glycopeptides and glycoproteins for biological investigation. *Annual Review of Biochemistry* **71**, 593-634 (2002).

56. Marcaurelle, L. A. et al. Chemical synthesis of lymphotactin: a glycosylated chemokine with a C-terminal mucin-like domain. *Chemistry (Weinheim an der Bergstrasse, Germany)* **7**, 1129-32 (2001).

57. Miller, J. S., Dudkin, V. Y., Lyon, G. J., Muir, T. W. & Danishefsky, S. J. Toward fully synthetic N-linked glycoproteins. *Angew. Chem., Int. Ed. Engl.* **42**, 431-434 (2003).

58. Hojo, H. et al. The first synthesis of peptide thioester carrying N-linked core pentasaccharide through modified Fmoc thioester preparation: synthesis of an N-glycosylated Ig domain of emmprin. *Tetrahedron Lett.* **44**, 2961-2964 (2003).

59. Tolbert Thomas, J., Franke, D. & Wong, C.-H. A new strategy for glycoprotein synthesis: ligation of synthetic glycopeptides with truncated proteins expressed in E. coli as TEV protease cleavable fusion protein. *Bioorganic & medicinal chemistry* **13**, 909-15 (2005).

60. Hackenberger Christian, P. R., Friel Claire, T., Radford Sheena, E. & Imperiali, B. Semisynthesis of a glycosylated Im7 analogue for protein folding studies. *J. Am. Chem. Soc.* **127**, 12882-9 (2005).

61. Durek, T. et al. Synthesis of fluorescently labeled mono- and diprenylated Rab7 GTPase. *J. Am. Chem. Soc.* **126**, 16368-16378 (2004).

62. Brunsveld, L. et al. Synthesis of functionalized Rab GTPases by a combination of solution- or solid-phase lipopeptide synthesis with expressed protein ligation. *Chem.-Eur. J.* **11**, 2756-2772 (2005).

63. Gottlieb, D., Grunwald, C., Nowak, C., Kuhlmann, J. & Waldmann, H. Intein-mediated in vitro synthesis of lipidated Ras proteins. *Chem. Commun.*, 260-262 (2006).

64. Grogan Michael, J., Kaizuka, Y., Conrad Rosemary, M., Groves Jay, T. & Bertozzi Carolyn, R. Synthesis of lipidated green fluorescent protein and its incorporation in supported lipid bilayers. *J. Am. Chem. Soc.* **127**, 14383-7 (2005).

65. Shogren-Knaak, M. A. & Peterson, C. L. Creating designer histones by native chemical ligation. *Methods Enzymol.* **375**, 62-76, 1 Plate (2004).

66. Shogren-Knaak, M. et al. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science (Washington, D. C.)* **311**, 844-7 (2006).

67. Scheibner, K. A., Zhang, Z. & Cole, P. A. Merging fluorescence resonance energy transfer and expressed protein ligation to analyze protein-protein interactions. *Anal. Biochem.* **317**, 226-232 (2003).

68. Maag, D. & Lorsch, J. R. Communication between eukaryotic translation initiation factors 1 and 1A on the yeast small ribosomal subunit. *J. Mol. Biol.* **330**, 917-924 (2003).

69. Muralidharan, V. et al. Domain-specific incorporation of noninvasive optical probes into recombinant proteins. *J. Am. Chem. Soc.* **126**, 14004-14012 (2004).

70. Becker Christian, F. W. et al. Total chemical synthesis of a functional interacting protein pair: the protooncogene H-Ras and the Ras-binding domain of its effector c-Raf1. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5075-80 (2003).

71. Yeo, D. S. Y. et al. Cell-permeable small molecule probes for site-specific labeling of proteins. *Chem. Commun.*, 2870-2871 (2003).

72. Wang, D. & Cole, P. A. Protein tyrosine kinase Csk-catalyzed phosphorylation of Src containing unnatural tyrosine analogues. *J. Am. Chem. Soc.* **123**, 8883-8886 (2001).

73. Berry, S. M., Ralle, M., Low, D. W., Blackburn, N. J. & Lu, Y. Probing the role of axial methionine in the blue copper center of azurin with unnatural amino acids. *J. Am. Chem. Soc.* **125**, 8760-8768 (2003).

74. Arnold, U. et al. Protein prosthesis: A semisynthetic enzyme with a β-peptide reverse turn. *J. Am. Chem. Soc.* **124**, 8522-8523 (2002).

75. Shen, K. & Cole, P. A. Conversion of a tyrosine kinase protein substrate to a high affinity ligand by ATP linkage. *J. Am. Chem. Soc.* **125**, 16172-16173 (2003).

76. Jantz, D. & Berg, J. M. Expanding the DNA-Recognition Repertoire for Zinc Finger Proteins beyond 20 Amino Acids. *J. Am. Chem. Soc.* **125**, 4960-4961 (2003).

77. Lu, W., Gong, D., Bar-Sagi, D. & Cole, P. A. Site-specific incorporation of a phosphotyrosine mimetic reveals a role for tyrosine phosphorylation of SHP-2 in cell signaling. *Mol. Cell* **8**, 759-769 (2001).

78. Lu, W., Shen, K. & Cole, P. A. Chemical dissection of the effects of tyrosine phosphorylation of SHP-2. *Biochemistry* **42**, 5461-5468 (2003).

79. Zhang, Z., Shen, K., Lu, W. & Cole, P. A. The role of C-terminal tyrosine phosphorylation in the regulation of SHP-1 explored via expressed protein ligation. *J. Biol. Chem.* **278**, 4668-4674 (2003).

80. Schwarzer, D., Zhang, Z., Zheng, W. & Cole, P. A. Negative regulation of a protein tyrosine phosphatase by tyrosine phosphorylation. *J. Am. Chem. Soc.* **128**, 4192-4193 (2006).

81. Zheng, W. et al. Cellular stabilization of the melatonin rhythm enzyme induced by nonhydrolyzable phosphonate incorporation. *Nat. Struct. Biol.* **10**, 1054-1057 (2003).

82. Zheng, W. et al. Cellular stability of serotonin N-acetyltransferase conferred by phosphonodifluoromethylene alanine (Pfa) substitution for Ser-205. *J. Biol. Chem.* **280**, 10462-10467 (2005).

83. Hahn, M. E. & Muir, T. W. Bioorganic chemistry: Photocontrol of Smad2, a multiphosphorylated cell-signaling protein, through caging of activating phosphoserines. *Angew. Chem., Int. Ed. Engl.* **43**, 5800-5803 (2004).

84. Grant, J. E. et al. The N terminus of GTPγS-activated transducin α-subunit interacts with the C terminus of the cGMP phosphodiesterase γ-subunit. *J. Biol. Chem.* **281**, 6194-6202 (2006).

85. Xu, R., Ayers, B., Cowburn, D. & Muir, T. W. Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 388-393 (1999).

86. Blaschke, U. K., Cotton, G. J. & Muir, T. W. Synthesis of Multi-Domain Proteins Using Expressed Protein Ligation: Strategies for Segmental Isotopic Labeling of Internal Regions. *Tetrahedron* **56**, 9461-9470 (2000).

87. Anderson, L. L., Marshall, G. R., Crocker, E., Smith, S. O. & Baranski, T. J. Motion of Carboxyl Terminus of Gα Is Restricted upon G Protein Activation: A solution NMR study using semisynthetic Gα subunits. *J. Biol. Chem.* **280**, 31019-31026 (2005).

88. Becker, C. F. W. et al. Incorporation of spin-labelled amino acids into proteins. *Magn. Reson. Chem.* **43**, S34-S39 (2005).

89. Lesaicherre, M.-L., Lue, R. Y. P., Chen, G. Y. J., Zhu, Q. & Yao, S. Q. Intein-mediated biotinylation of proteins and its application in a protein microarray. *J. Am. Chem. Soc.* **124**, 8768-8769 (2002).

90. Richter, M. P. O., Holland-Nell, K. & Beck-Sickinger, A. G. Site specific biotinylation of the human aldo/keto reductase AKR1A1 for immobilization. *Tetrahedron* **60**, 7507-7513 (2004).

91. Lovrinovic, M. et al. Synthesis of protein-nucleic acid conjugates by expressed protein ligation. *Chem. Commun.*, 822-823 (2003).

92. Camarero, J. A., Cheung, C. L., Coleman, M. A. & De Yoreo, J. J. Chemoselective attachment of biologically active proteins to surfaces by native chemical ligation. *Materials Research Society Symposium Proceedings* **EXS-1**, 237-239 (2004).

93. Girish, A. et al. Site-specific immobilization of proteins in a microarray using intein-mediated protein splicing. *Bioorg. Med. Chem. Lett.* **15**, 2447-2451 (2005).

Chapter 2

Synthesis and characterization of caged phosphopeptides

Introduction

Protein phosphorylation is a central regulatory mechanism in signal transduction pathways and cell cycle regulation. An estimated one third of mammalian proteins undergo phosphorylation, and this post-translational modification can modulate enzymatic activity and protein-protein binding interactions.[3] Tools for the study of phosphorylation events in "real time" are essential for dissecting the dynamic role of phosphorylated residues in complex signaling networks. Increased understanding of signaling networks is of particular interest in cellular migration and related biological and medicinal studies.[4] Current methods for investigating phosphorylation, including gene knockout, RNA interference, and point mutation strategies, can be used to verify the essentiality of a given phosphorylation site, but are limited when probing the temporal role of phosphorylation in signaling pathways. As a complement to existing genetic techniques, we have developed caged phosphopeptides, which release a concentration burst of the corresponding phosphorylated peptides with temporal and spatial control following irradiation with long wavelength UV light (**Fig. 2.1**). These photolabile precursors allow investigators to probe the downstream effects of kinase-mediated phosphorylation in real time without altering upstream genetics.

The caging, or photolabile protection, of biomolecules enables the controlled release of an activated species. A caged substrate contains a protecting group that masks an essential functionality and that can be removed by photolysis to reveal a biologically active molecule.[5] For use in living systems, it is essential that both the caged compound and photolysis by-product be inert in the cell system. In addition, a useful caging group should be stable to hydrolysis and enzymatic cleavage under working conditions and have a fast rate of uncaging compared to the biological event under investigation.[6] The

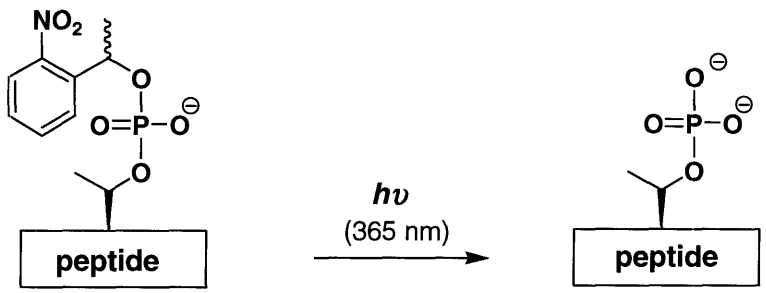


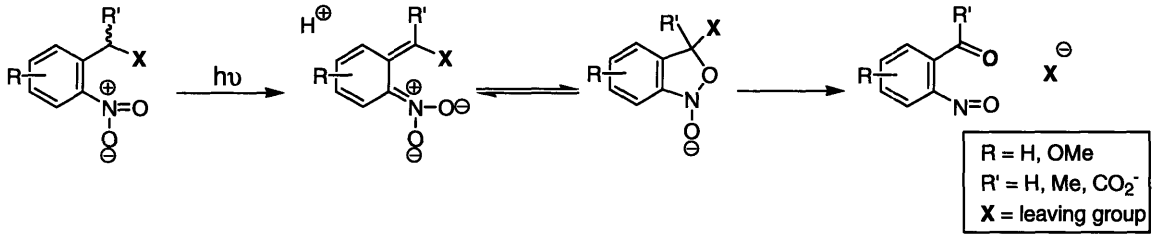
Figure 2.1. A 2-nitrophenylethyl-caged phosphopeptide. Irradiation with long wavelength UV light releases the corresponding phosphorylated peptide.

caging group must also release the effector molecule with reasonable quantum efficiency at a wavelength that does not cause radiative cellular damage. Derivatives of the *o*-nitrobenzyl group have been commonly employed to cage biologically active molecules, including calcium,[7] ATP, and peptides.[8,9] The *o*-nitrobenzyl group can be uncaged on a microsecond to millisecond timescale with long-wavelength UV light ($\lambda > 350$ nm) that is non-damaging to cells. A proposed mechanism for the photolysis of *o*-nitrobenzyl caged molecules is shown in **Scheme 2.1**.[10] For our syntheses of caged phosphopeptides for the study of signal transduction, the *o*-nitrobenzyl derivative 2-nitrophenylethyl was selected because it retains the desirable qualities of the *o*-nitrobenzyl moiety and releases a cellularly inert photo-byproduct. Irradiation of 2-nitropheylethyl-protected molecules releases 2-nitrosoacetophenone, which is significantly less reactive in the cellular environment than the corresponding aldehyde produced by the *o*-nitrobenzyl group or other derivatives lacking substitution at the benzylic position.[11] In addition, glutathione, which is naturally present in cells, reacts rapidly with nitroso ketones, preventing unwanted side reactions with the 2-nitrosoacetophenone.[10,11] The 2-nitrophenylethyl group has a moderately high extinction coefficient (ε_{max} = 5,700 $M^{-1}cm^{-1}$) with a maximum absorption at 259 nm.[12] The group has a broad absorption spectrum that enables high absorption even of long wavelength UV light. Therefore, all uncaging described in this thesis was carried out on a DNA transilluminator with emission centered at 365 nm.

Herein are described two synthetic methods for assembling 2-nitrophenylethyl caged phosphopeptides: 1) an interassembly approach incorporated into Fmoc-based

solid phase peptide synthesis (SPPS), and 2) a building block approach. Each method relies on a novel phosphitylating reagent to install the caging moiety. The interassembly approach enables access to caged phosphopeptides in laboratories set up for performing organic chemistry on solid support and works best for peptides without oxidation-sensitive residues *C*-terminal to the phosphitylated amino acid. The building block approach is accessible to any laboratory equipped for peptide synthesis and can be used to generate any caged phosphopeptide within the limits of SPPS. A biologically relevant caged phosphothreonine peptide, Ac-MARHFD(cpT)YLIRR-NH_2 (**cpChk2**) ('cp' = caged phospho-) was synthesized via these methods as a potential target for biological studies. The peptide was characterized by analytical high performance liquid chromatography (HPLC) and electrospray ionization mass spectrometry (ES-MS), and the quantum yield of uncaging was measured and found to be comparable to literature precedent for *o*-nitrobenzyl-caged molecules.

Scheme 2.1. Photolysis of molecules masked by *o*-nitrobenzyl-derived caging groups.[10]

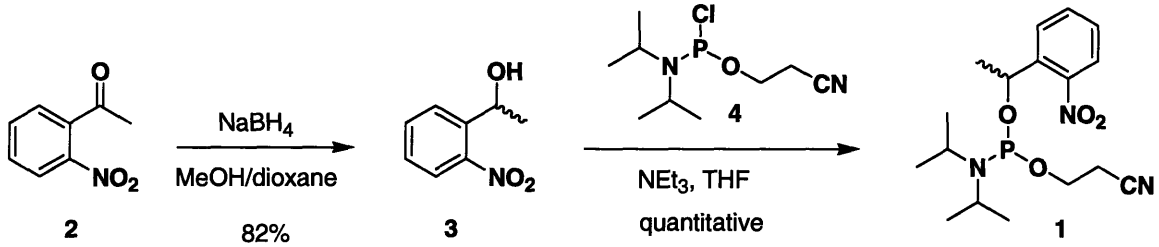


Results and Discussion

Both the interassembly and building block strategies utilize a phophitylating agent for the introduction of a 2-nitrophenylethyl caged phosphite, which is subsequently oxidized to the corresponding phosphate. The phosphitylating reagent, *O*-1-(2-nitrophenyl)ethyl-*O*'-*β*-cyanoethyl-*N,N*-diisopropylphosphoramidite (**1**) was synthesized as shown in **Scheme 2.2.** First 2-nitroacetophenone (**2**) was reduced with sodium borohydride to a racemic mixture of 1(2-nitrophenyl)-ethanol (**3**). Displacement of the chloride by the alcohol (**3**) in the DNA synthesis reagent 2-cyanoethyl diisopropylchlorophosphoramidite (**4**) [13] afforded phosphoramidite **1**. Importantly the 2-

nitrophenylethyl caging group is stable to the acidic and basic conditions required for incorporation into peptides in the Fmoc-based SPPS approaches described below.

Scheme 2.2. Synthesis of phosphitylating agent **1**.

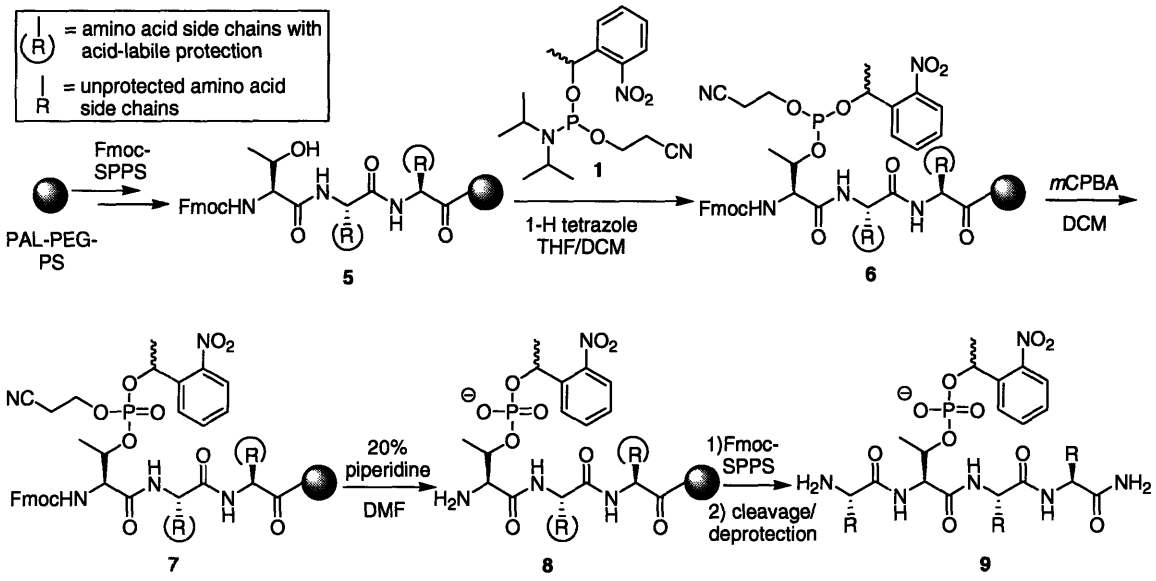


2-1. Interassembly approach

The construction of caged phosphopeptides via the interassembly approach is illustrated with a threonine-containing peptide in **Scheme 2.3**. The "interassembly" designation refers to the introduction of a modification onto a growing peptide chain. Accordingly, peptide **5** was synthesized on a PAL-PEG-PS solid support by standard Fmoc-based SPPS up to the threonine residue to be modified, which was introduced without side chain protection. The free threonine residue was phosphitylated with **1** activated by 1*H*-tetrazole in 1:1 dichloromethane (DCM): tetrahydrofuran (THF), yielding the phosphite peptide **6** on resin. The trivalent phosphorus species was then oxidized to the bisprotected phosphate **7** with *m*-chloroperoxybenzoic acid (*m*CPBA). Subsequent treatment with 20% piperidine in *N,N*-dimethylformamide (DMF) for Fmoc removal in chain elongation simultaneously removed the cyanoethyl group to afford the monoprotected caged phosphate **8**. While phosphotriesters are highly susceptible to *β*–elimination in the basic Fmoc-removal conditions, phosphodiesters are stable to deprotection conditions and adequately protect the phosphate moiety from side reactions during SPPS.[14] Following peptide completion, the peptide was cleaved with trifluoroacetic acid (TFA) in the presence of scavengers to yield tetrapeptide **9**. Caged phosphoserine and caged phosphotyrosine-containing peptides were accessed using similar methods.[1]

As a representative caged phosphothreonine peptide, Ac-MARHFD(cpT)YLIRR-NH_2 (**cpChk2**) was synthesized via the interassembly approach (**Fig. 2.2**) and analyzed at

Scheme 2.3. Synthesis of a caged phosphothreonine peptide (**9**) via the interassembly approach.



key intermediates by analytical HPLC and ES-MS. The corresponding uncaged peptide would be an antagonist of Chk2, a homologue of the human Rad53p checkpoint kinase, which contains a single forkhead associated (FHA) domain.[15] FHA domains are phosphothreonine-binding modules that can be found in various transcriptional control proteins, kinases activated by DNA damage, phosphatases, and cell cycle checkpoint proteins. Chk2 is known to cause cell cycle arrest and is one of several FHA domain-containing proteins that may play a role in protecting against cancer. It has been suggested that mutations in Chk2 are involved in the formation of tumors.[15,16]

In the synthesis of **cpChk2**, the synthetic intermediate with a cyanoethyl-protected phophosphorylated threonine (**7**) appeared as two separate peaks in the HPLC

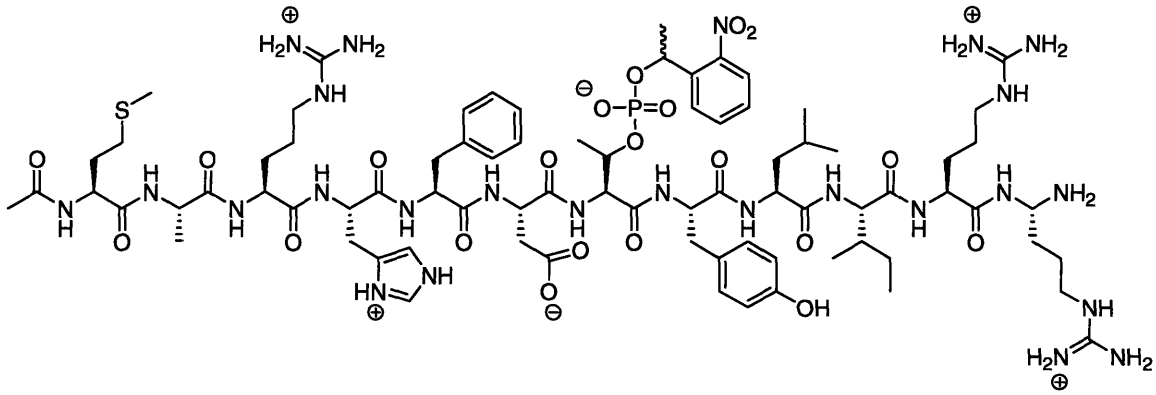


Figure 2.2. Caged phosphopeptide **cpChk2**, Ac-MARHFD(cpT)YLIRR-NH_2.

chromatogram. The dual peaks likely represent diastereomers resulting from the adjacent chiral centers at the phosphorus and at the β carbon on the threonine, since the two peaks showed identical masses in mass spectrometry analysis. After deprotection of the cyanoethyl group with piperidine, resulting in the loss of chirality at phosphorus, the HPLC chromatogram showed a single peak for the peptide. Due to the potential for methionine oxidation, the **cpChk2** peptide was dissolved in deoxygenated water and stored at -20 °C.

With the interassembly approach, residues sensitive to oxidation should only be included in a caged phosphopeptide sequence *N*-terminal to the modified residue, such as in **cpChk2**. Oxidation of the phosphite intermediate (**6**) could potentially also affect susceptible residues, namely tryptophan and methionine, present in the peptide (**Fig. 2.3**).

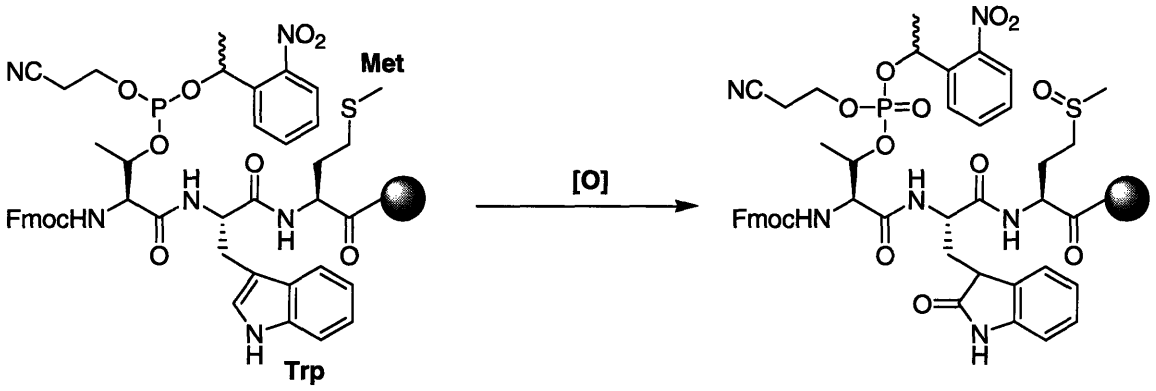


Figure 2.3. Possible oxidation of methionine and tryptophan resides during oxidation of a phosphite intermediate to the corresponding phosphate in the interasssembly approach.

2-2. Building block approach

The synthesis of *N*-α-Fmoc-protected 2-nitrophenylethyl-caged phosphothreonine (**10**), phosphoserine (**11**), and phosphotyrosine (**12**) building blocks facilitate the straightforward assembly of any caged phosphopeptide (**Fig. 2.4**). Specifically the building block approach overcomes the two restrictions of the interassembly strategy; undesired oxidation of tryptophan or methionine residues, and limited accessibility of the caged phosphopeptides to laboratories not equipped for organic chemistry manipulations on solid support. With the appropriate building blocks, any researcher can gain access to the desired caged phosphopeptides through standard SPPS or through a commercial peptide synthesis facility.

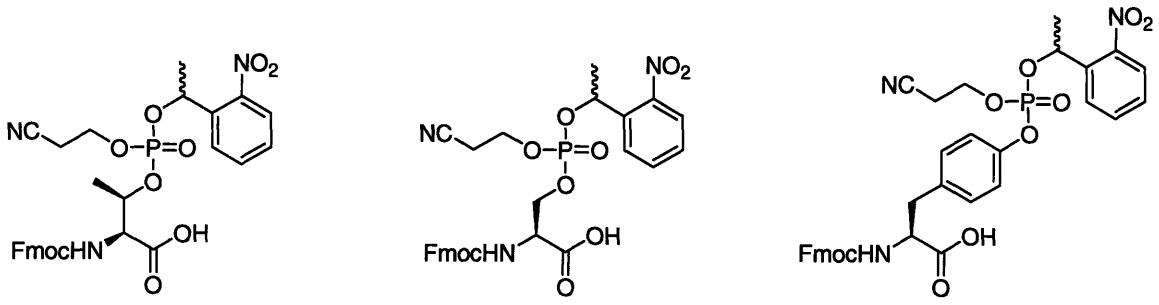


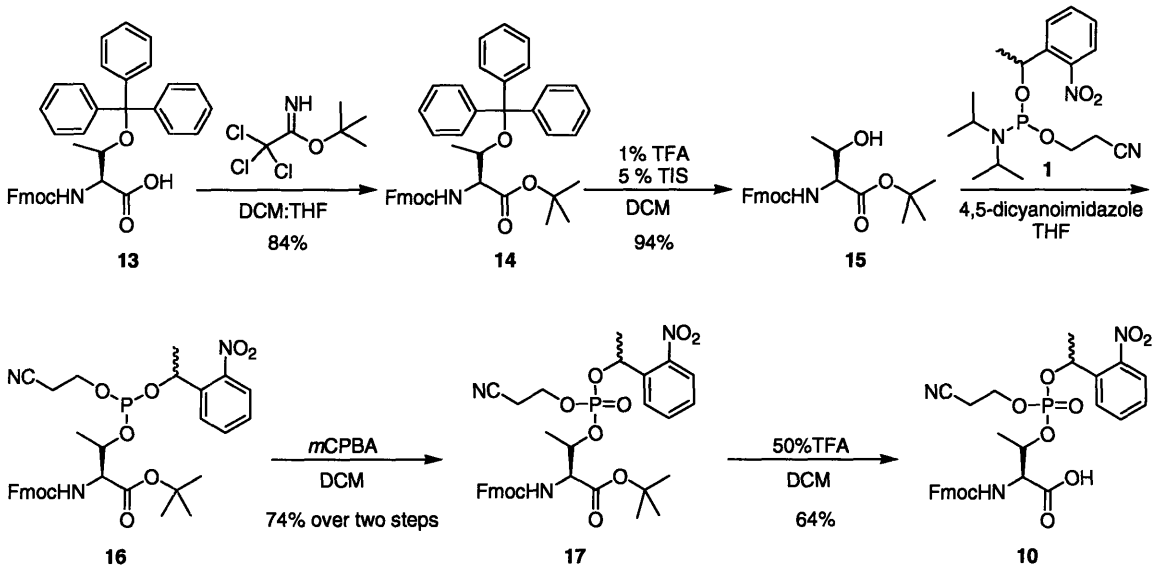
Figure 2.4. Structure of caged phosphoamino acid building blocks for Fmoc-based SPPS: **10**, *N*-α-Fmoc-phospho(1-nitro-phenylethyl-2-cyanoethyl)-L-threonine; **11**, *N*-α-Fmoc-phospho(1-nitro-phenylethyl-2-cyanoethyl)-L-serine; **12**, *N*-α-Fmoc-phospho(1-nitro-phenylethyl-2-cyanoethyl)-L-tyrosine.

The strategy for constructing caged phosphoamino acids is similar to that of the interassembly approach and utilizes the identical phosphitylating reagent (**1**) for the P-O bond construction. The general synthesis of the caged building blocks is depicted for caged phosphothreonine (**10**) in **Scheme 2.4**. Commercially available *N*-α-Fmoc-hydroxytrityl-L-serine (**13**) was treated with *tert*-butyl *2,2,2*-trichloroacetimidate to yield the *N*-α-Fmoc *tert*-butyl ester **14**. The trityl protecting group was then removed under mild acidic conditions to reveal **15**. The *N*-α-Fmoc *tert*-butyl ester **15** was phosphitylated with **1** activated by 4,5-dicyanoimidazole to afford **16**. The phosphite was then oxidized without prior purification to the bisprotected phosphate **17** with *m*CPBA. Subsequent removal of the *tert*-butyl protection with 50% TFA in DCM at room temperature afforded the caged phosphothreonine building block **10**, which was stored at -20 °C or used directly in SPPS. Importantly, the final deprotection reaction required temperatures of 23 °C or warmer; deprotection failed at lower "room" temperatures. Caged phosphoserine (**11**) and caged phosphotyrosine (**12**) were similarly synthesized.[2]

During incorporation into target peptides, the building blocks retain the β-cyanoethyl protection. Following the peptide coupling step, the base-mediated Fmoc deprotection concurrently removes the β-cyanoethyl group. The resulting monoprotected phosphates are ideal for Fmoc-based SPPS because β-elimination side reactions of phosphothreonine and phosphoserine residues are suppressed, as discussed in **Section 2-**

1.[14] The building blocks have been used in SPPS for the successful syntheses of a variety of caged phosphorylated peptides.

Scheme 2.4. Synthesis of *N*-α-Fmoc-phospho(1-nitro-phenylethyl-2-cyanoethyl)-L-threonine, **10**, for Fmoc-based SPPS.



2-3. Quantum yield calculation

The quantum yield of photolysis, Φ, a measure of uncaging efficiency, was determined for **cpChk2** by comparing the percent loss of the caged peptide after photolysis with that of a standard caged sample with known Φ, as quantified by analytical HPLC. Caged phosphate (2-nitrophenylethyl-phosphate, Φ = 0.54),[17] was used as the standard for these calculations. Solutions of **cpChk2** and the caged phosphate standard were prepared in water at pH 7.1 with 5 mM DTT and 0.46 mM inosine. The DTT was added to prevent any back reaction of the photo-byproduct, 2-nitrosoacetophenone, with the liberated phosphopeptides. The inosine provided a photochemically inert internal standard to allow for quantification by HPLC analysis of the relative peak areas of **cpChk2** and caged phosphate before and after photolysis.

The solutions of **cpChk2** and caged phosphate were each irradiated for 15 seconds with light centered at λ = 365 nm and at an intensity of 7,330 μW/cm^2 in glass vessels with path lengths of 1 mm. Analysis by HPLC was performed prior to (t = 0 s) and immediately after (t = 15 s) irradiation. According to literature precedent, the A_{350}

values of the caged peptide and the caged phosphate were measured and used to calculate the quantum yield of uncaging with the following equation:[17]

$$\Phi_{sample} = \Phi_{standard} \times (\%\Delta sample / \%\Delta standard) \times (A_{350}standard/A_{350}sample),$$

where $\%\Delta$ is the percent loss measured by HPLC and A_{350} is the absorbance of the caged compound at 350 nm, in water at pH 7.1. The experiment was performed in triplicate and the quantum yield for **cpChk2** was found to be 0.33, which is within the range of reported literature values for *o*-nitrobenzyl-caged molecules. The Φ of caged phosphopeptides appears to be peptide dependent, as Φ values determined for several other caged phosphopeptides were found to range between 0.2 and 0.4.[1]

Conclusion

The synthesis of caged phosphopeptides has been described using an interassembly and a building block approach. While both approaches facilitate access to a wide range of peptides, the building block approach is the most versatile, enabling the synthesis of peptides with oxidation-sensitive residues anywhere in the sequence. The caged phosphothreonine peptide **cpChk2** was synthesized and the quantum yield of uncaging was determined. Caged phosphopeptides synthesized using the approaches described here have been successfully applied to in vitro and in vivo experiments to study kinase-mediated signal transduction in cellular migration and cell cycle control.[18,19]

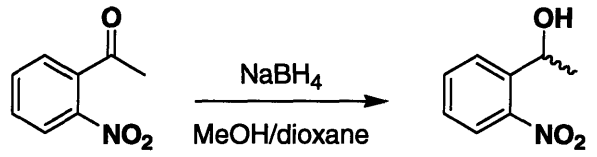
Methods

General Methods. All peptide synthesis reagents were purchased from Novabiochem and all other chemicals were purchased from Aldrich. Dry DCM was distilled from calcium hydride and dry THF was distilled from sodium/benzophenone. HPLC was performed on a Waters 400 or 600 system with a Waters dual wavelength absorbance detector reading at 228 and 280 nm. For C_{18} analytical HPLC a Beckman Ultrasphere ODS, 5 μm, 150 x 4.6 mm column was used and for C_{18} preparatory HPLC a YMC-Pac Pro, 5 μm, 250 x 20 mm column was used. The standard HPLC separation gradient was 93:7 to 0:100 water:MeCN with 0.1% TFA over 30 minutes. ES-MS was performed on a PerSeptive BioSystems Mariner Biospectrometry Workstation. NMR spectra were acquired on a

BrukerAdvance (DPX) 400 MHz instrument, or a Varian Mercury 300 MHz instrument. Reported chemical shifts are in ppm relative to a standard (TMS for ^{1}H or chloroform for ^{13}C).

General Peptide Synthesis. Peptides were prepared by standard Fmoc SPPS on Fmoc-PAL-PEG-PS resin. For example, for 0.04 mmol of peptide, 200 mg of Fmoc-PAL-PEG-PS resin (0.2 mmol/g loading) were added to a peptide synthesis vessel and swollen for five minutes in 5 mL of DCM, then in 5 mL of DMF. The resin was deblocked with three five-minute rinses with 5 mL of piperidine, and then rinsed for five one-minute rinses with 5 mL DMF. An Fmoc-protected amino acid was then coupled as follows. For each coupling, 80 μmol of Fmoc-protected amino acid and 80 μmol of benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP) were dissolved in 5 mL of DMF and then added to the deblocked resin. The couplings were initiated by the addition of 160 μmol of diisopropylethylamine (DIEA), and then agitated for at least 30 minutes at room temperature. The resin was rinsed twice with one-minute washes of 5 mL of DMF, then twice with 5 mL washes of DCM. Following each coupling, several resin beads were tested for free amines with trinitrobenzene sulfonic acid (TNBS). After a negative test, the amino acid was deblocked as described above and chain elongation was continued. Peptides were cleaved from the resin under a variety of conditions, depending on the peptide composition. Cleaved peptides were triturated three times with cold diethylether, and purified by reverse phase HPLC. The identities of the peptides were confirmed by ES-MS (turbo ion source). The concentration of caged phosphopeptide solutions was determined by UV absorption based on the nitrophenyl group (ε_{259} = 5700 $M^{-1}cm^{-1}$ in MeOH). The concentration of peptide solutions with tyrosine or tryptophan residues was determined by UV absorption based on those residues (ε_{280} = 5500(#Trp) + 1490(#Tyr) $M^{-1}cm^{-1}$ in water).[20] The concentration of all other peptide solutions was determined by quantitative amino acid analysis.

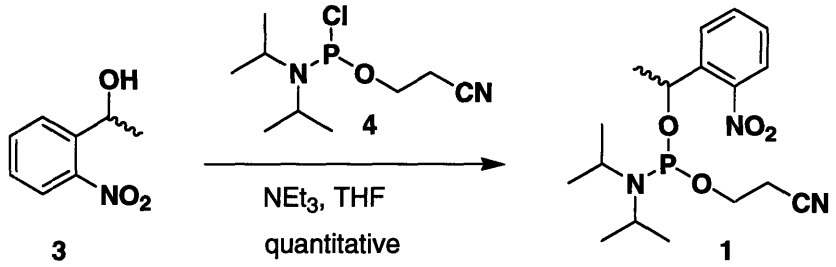
1(2-nitrophenyl)-ethanol (**3**)



2 82% 3

To a stirring solution of 2-nitroacetophenone (7.0 g, 42.3 mmol) in 130 mL of dioxane:methanol (2:3 by volume), sodium borohydride (4.8 g, 126.9 mmol) was added over a period of 20 minutes and stirred in a water bath at 0 °C. The solution was allowed to stir for an additional 2.5 hours at room temperature and the progress was monitored by TLC in chloroform. The reaction was quenched with water (150 mL) and then extracted into chloroform (3 x 100 mL). The organic fractions were dried over magnesium sulfate ($MgSO_4$) and concentrated under reduced pressure. The product was further dried *in vacuo* overnight to yield an orange-yellow oil, 5.81 g (82.0%). ^1H NMR (300 MHz, $CDCl_3$) δ ppm: 7.9 (d, J = 8.0 Hz, 1 H), 7.7 (t, J = 7.2 Hz, 1 H), 7.4 (t, J = 7.6 Hz, 1 H), 5.4 (q, J = 6.4 Hz, 12.4 Hz, 1 H), 2.4 (s, 1 H), 1.6 (d, J = 6.4 Hz, 3 H). ^{13}C NMR (126 MHz, $CDCl_3$) δ ppm: 148.1, 141.5, 134.1, 128.5, 128.0, 124.7, 65.9, 24.7.

O-*1-(2-nitrophenyl)ethyl*-O'-*β-cyanoethyl*-N,N-*diisopropylphosphoramidite* (**1**).



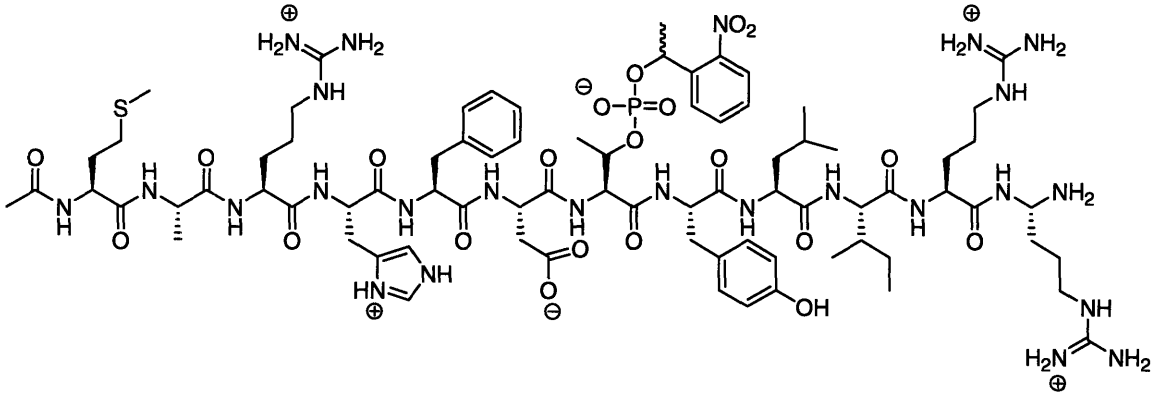
3 1

In the dark, 2-cyanoethyl diisopropylchlorophosphoramidite (470 μL, 2.11 mmol) was added to a stirring solution of nitrophenylethanol (290 mg, 1.76 mmol) and triethylamine (590 μL, 4.23 mmol) in anhydrous DCM (10 mL). The reaction mixture was stirred overnight at room termperature and the reaction progress was monitored by TLC (7:3:4 methanol : chloroform : water). The solution was washed with 10% $NaHCO_3$. The organic layer was dried over $MgSO_4$, concentrated under reduced pressure, then further dried overnight *in vacuo* to yield a deep-yellow oil, 645 mg (quantitative). ^1H NMR (300 MHz, $CDCl_3$) δ ppm: 7.9 (m, 2 H), 7.7 (m, 1 H), 7.4 (m, 1

H), 5.6 (m, 1 H), 3.9 (m, 2 H), 3.7 (m, 2 H), 2.5 (m, 2H), 1.6 (dm, J = 2.4 Hz, 6.8 Hz, 3 H), 1.2 (m, 12H). ^{31}P NMR (121 MHz, $CDCl_3$) δ: 148.2 (d, J = 45.6 Hz).

Caged phosphothreonine peptides (interassembly approach). Peptides were built up on a PAL-PEG-PS solid support by standard Fmoc SPPS up to the threonine residue to be modified, which was introduced without side-chain protection, and the resulting peptide on resin was pumped dry overnight. All subsequent steps were carried out in the dark. The threonine was coupled with *O*-1-(2-nitrophenyl)ethyl-*O*'-β-cyanoethyl-*N,N*-diisopropylphosphoramidite (5 equivalents) with 1 *H*-tetrazole (10 equivalents) in 1:1 DCM:THF for 1 hour. The reaction was quenched with 10% $NaHCO_3$, and the resin was washed with 1:1 DCM:THF, followed by DCM. The peptide on resin was oxidized with *m*CPBA (2 equivalents) in DCM for 1 hour. The reaction was quenched with 10% $NaHCO_3$, and the resin was washed with DCM. Subsequent treatment with 20% piperidine in DMF for Fmoc deprotection simultaneously removed the the β-cyanoethyl group. Peptide completion was carried out with standard SPPS protocols. The synthesis of **cpChk** is detailed as an example below:

***cpChk2**: Ac-MARHFD-phospho(nitrophenylethyl)-threonyl-YLIRR-CONH$_2$.*



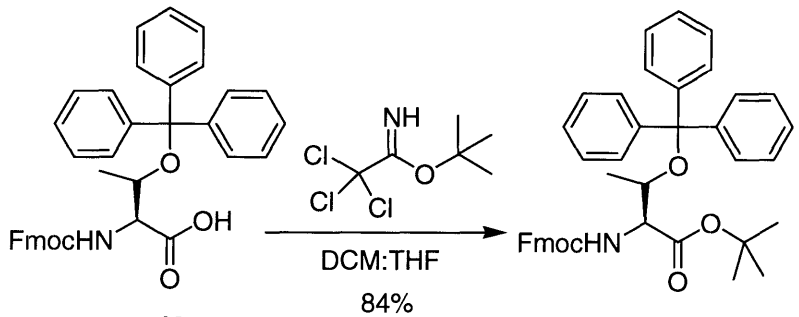
Fmoc-TYLIRR-(PAL-PEG-PS). Fmoc-TYLIRR-$CONH_2$ was prepared using standard Fmoc SPPS as described above on PAL-PEG-PS resin. A test cleavage was run. Reverse phase HPLC (t_R = 27.4 min). Exact mass calcd for $C_{52}H_{75}N_{13}O_{10}$, 1042.3; found by MS(ESI), 521.8 $[M2H]^{2+}$.

Fmoc-phosphi-(*O*-1-(2-nitrophenyl)ethyl-*O'*-*β*-cyanoethyl)threonyl-YLIRR-(PAL-PEG-PS). Fmoc-TYLIRR-(PAL-PEG-PS) (114 mg resin, 25 μmol) was pumped dry *in vacuo* overnight. *O*-1-(2-nitrophenyl)ethyl-*O'*-*β*-cyanoethyl-*N,N*-diisopropylphosphoramidite (92 mg, 250 μmol) was added to the resin and the system was flushed with argon. The resin was swollen in anhydrous DCM for 15 minutes. The system was charged with 1*H*-tetrazole (17.5 mg, 250 μmol), and then flushed again with argon. The reaction was stirred for overnight in the dark, and saturated $NaHCO_3$ (4 mL) was added. The resin was rinsed with 1:1 DCM:THF (2 x 5 mL) and DCM (2 x 5mL).

Fmoc-phospho-(*O*-1-(2-nitrophenyl)ethyl-*O'*-*β*-cyanoethyl)threonyl-YLIRR-(PAL-PEG-PS). Fmoc-phosphi-(*O*-1-(2-nitrophenyl)ethyl-*O'*-*β*-cyanoethyl)threonyl-YLIRR-(PAL-PEG-PS) (114 mg resin, 25 μmol) was swollen with DCM in a peptide synthesis vessel. To this was added *m*CPBA (8.6 mg, 50 μmol) in DCM (2.8 mL) and the mixture was agitated in the dark for 1 hour. Saturated $NaHCO_3$ (6 mL) was added to the swirling solution. After 5 minutes the resin was rinsed with DCM (4 x 5 mL). A test cleavage was run. Two peaks observed by reverse phase HPLC (t_R = 28.6 min and 28.8 min). Exact mass calcd for $C_{63}H_{85}N_{15}O_{15}P$, 1324.5; found by MS(ESI), 662.9 $[M2H]^{2+}$. Identical mass found for the two HPLC peaks.

Ac-MARHFD-phospho-(1-nitrophenylethyl)-threonyl-YLIRR-CONH2. Sequence completion was carried out using standard Fmoc SPPS in the dark. The *N*-terminus was acetyl capped by coupling with acetic anhydride (26 μL, 0.28 mmol) and pyridine (23 μL, 0.28 mmol) in DMF (5 mL) for 30 min. The peptide was cleaved from the resin in 92.5% TFA, 2.5% water, 2.5% TIS, and 2.5% EDT. Reverse phase HPLC (t_R = 19.9 min). Exact mass calcd for $C_{83}H_{124}N_{26}O_{22}PS$, 1849.0; found by MS(ESI), 662.9 $[M2H]^{2+}$.

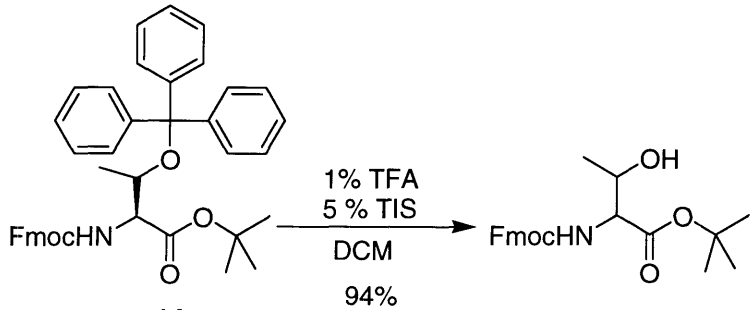
N^{α}-*Fmoc-hydroxyltrityl-L-threonine* tert-*butyl ester* (**14**)



13 **14**

To N^{α}-Fmoc-hydroxyltrityl-L-threonine (1.89 g, 3.24 mmol) in 4:1 DCM:THF (16 mL) was added *tert*-butyl *2,2,2*-trichloroacetimidate (2.3 mL, 12.95 mmol). The reaction was stirred for 2 hours at room temperature under argon and reaction progress was monitored by TLC (1:1 EtOAc:hexanes). The reaction mixture was concentrated, then redissolved into EtOAc (30 mL). The solution was washed with 10% $NaHCO_3$ (2 x 60 mL) and saturated sodium chloride (30 mL). The organic layer was dried over $MgSO_4$, filtered, and concentrated. The crude product was purified by flash chromatography on a short plug of basic alumina (1:1 hexanes/EtOAc, $R_f = 0.66$) to give the product (1.74 g) in 84% yield. ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 7.88 (d, J = 7.5 Hz, 2H), 7.81 (d, J = 7.3 Hz, 1H), 7.78 (d, J = 7.5 Hz, 1H), 7.61-7.34 (m, 19H), 5.98 (d, J = 9.7 Hz, 1H), 4.56 (m, 2H), 4.43 (m, 1H), 4.34 (m, 1H), 4.02 (m, 1H), 1.47 (s, 9H), 1.09 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ ppm: 169.9, 156.6, 144.6, 144.0, 143.8, 141.3, 129.0, 127.7, 127.6, 127.2, 127.1, 125.1, 125.2, 120.98, 119.96, 86.3, 81.9, 71.4, 67.1, 60.1, 47.2, 27.9, 18.6. ESI-MS: $[MNa]^+$ 646.1554 (obsd), 646.1561 (calcd).

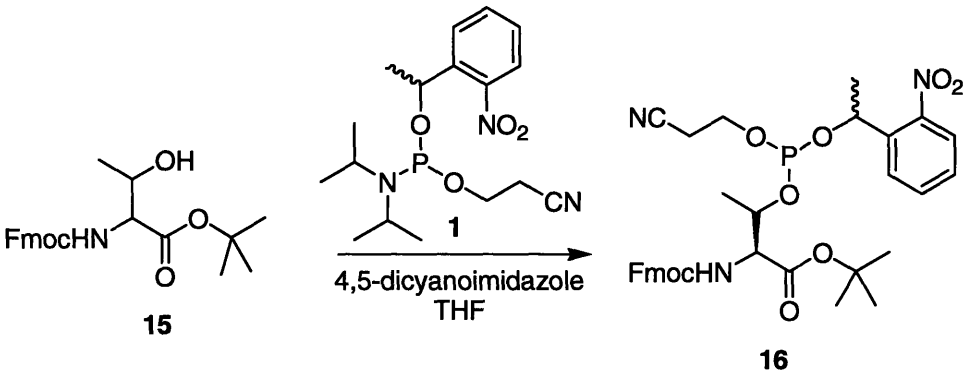
N^{α}-*Fmoc-L-threonine* tert-*butyl ester* (**15**)



14 **15**

N^{α}-Fmoc-hydroxyltrityl-L-threonine *tert*-butyl ester (3.66 g, 5.72 mmol) was dissolved in 150 mL of 1% TFA and 5% TIS in DCM. The reaction mixture was agitated at room temperature for 30 min. The solution was diluted with DCM (100 mL) and washed with 10% $NaHCO_3$ (2 x 120 mL) and concentrated sodium chloride (120 mL). The organic layer was dried over $MgSO_4$, filtered, and concentrated. The crude product was purified by silica gel flash chromatography (hexanes, then 1:1 hexanes/EtOAc, R_f = 0.48) to give the product (2.14 g) in 94% yield. ^1H NMR (300 MHz, $CDCl_3$) δ ppm: 7.75 (d, J = 7.4 Hz, 2H), 7.63 (d, J = 7.4 Hz, 2H), 7.39 (t, J = 7.2 Hz, 2H), 7.30 (t, J = 7.4 Hz, 2H), 5.99 (d, J = 9.1 Hz, 1 H), 4.42 (m, 2H), 4.35 (dd, J = 2.2 Hz and 6.3 Hz, 1H), 4.30 (m, 1H), 4.24 (t, J = 7.2 Hz, 1H), 3.07 (bs, 1H), 1.52 (s, 9H), 1.29 (d, J = 6.6 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ ppm: 170.4, 157.0, 143.9, 143.7, 141.3, 127.7, 127.1, 125.2, 120.0, 82.5, 68.2, 67.2, 59.8, 47.1, 28.0, 20.0. ESI-MS: $[MNa]^+$ 420.1770 (obsd), 420.1781 (calcd).

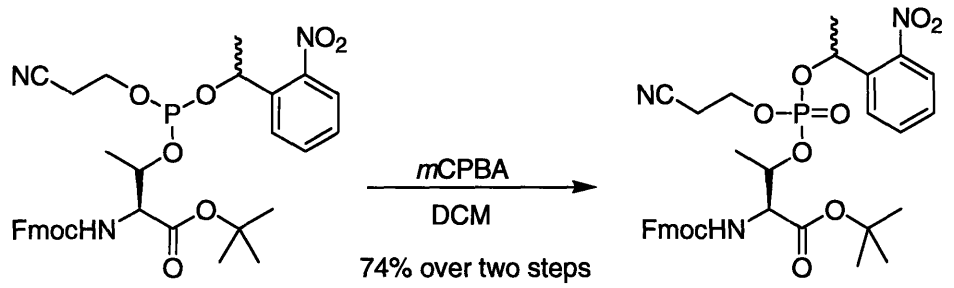
N^{α}-*Fmoc-phosphi(1-nitrophenylethyl-2-cyanoethyl)-L-threonine* tert-*butyl ester* (**16**)



In a round-bottom flask under argon was dissolved N-α-Fmoc-L-threonine *tert*-butyl ester (1.35 g, 3.39 mmol) in anhydrous THF (16 mL). In a pear-shaped flask under argon were dissolved *O*-1-(2-nitrophenyl)-ethyl-*O*'-*β*-cyanoethyl-*N*,*N*-diisopropylphosphoramidite (3.74 g, 10.2 mmol) and 4,5-dicyanoimidazole (1.20 g, 10.20 mmol) in anhydrous THF (16 mL), and the solution was mixed for several minutes in the dark. The phosphoramidite solution was added into the stirring amino acid solution via cannula under argon positive pressure and allowed to stir at room temperature in the dark overnight, under argon. The reaction was judged complete by disappearance of the

starting material and appearance of the product. The crude mixture was concentrated under reduced pressure and redissolved in EtOAc (100 mL). The solution was washed with 10% $NaHCO_3$ (2 × 100 mL) and then brine (100 mL). The crude material was dried over Na_2SO_4, filtered, concentrated under reduced pressure, and used immediately in the following reaction.

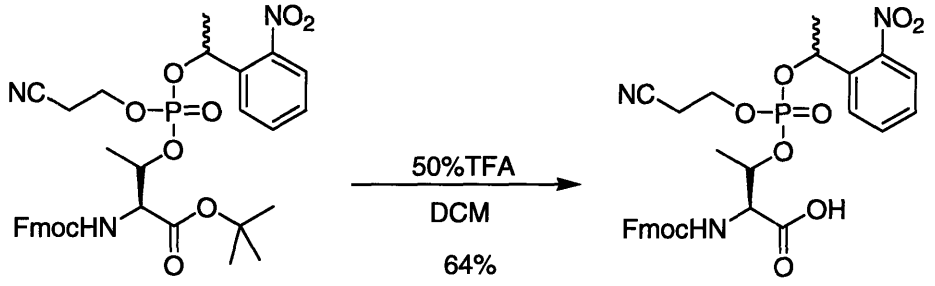
N-*α-Fmoc-phospho(1-nitrophenylethyl-2-cyanoethyl)-L-threonine* tert-*butyl ester* (**17**)



16 **17**

N-α-Fmoc-phosphi(1-nitrophenylethyl-2-cyanoethyl)-L-threonine *tert*-butyl ester was dissolved in DCM (130 mL). To this solution was added *m*CPBA (1.17 g, 6.78 mmol), and the resulting mixture was agitated at room temperature in the dark for 1 hour. The solution was washed with 10% $NaHCO_3$ (2 x 120 mL), and the organic layer was dried over $MgSO_4$, filtered, and concentrated. The product was purified by silica gel flash chromatography (1:3 hexanes/EtOAc, $R_f = 0.38$) to give the product (1.74 g) in 74% yield over two steps. ^1H NMR (300 MHz, $CDCl_3$) δ ppm: 7.96 (m, 1H), 7.80-7.60 (m, 6H), 7.53-7.33 (m, 5H), 6.12 (m, 1H), 5.45 (m, 1H), 5.03 (m, 2H), 4.44 (m, 2H), 4.24 (m, 3 H), 2.69 (m, 2H), 1.75 (m, 3H), 1.47 (m, 9H), 1.34 (m, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) □ppm: 168.5, 168.2, 156.8, 147.1, 143.8, 141.5, 136.9, 134.1, 129.4, 128.0, 127.9, 127.8, 127.3, 125.3, 124.8, 120.2, 116.4, 83.6, 73.9, 67.5, 67.4, 62.5, 58.9, 47.3, 28.1, 24.4, 19.8, 18.4. ^{31}P NMR (121.5 MHz, $CDCl_3$) δ ppm: -3.30, -3.20. ESI-MS: $[MNa]^+$ 702.2194 (obsd), 702.2187 (calcd).

N-α-*Fmoc-phospho(1-nitrophenylethyl-2-cyanoethyl)-L-threonine* (**10**)



17 **10**

N-α-Fmoc-phospho(1-nitrophenylethyl-2-cyanoethyl)-L-threonine *tert*-butyl ester (268 mg, 385 μmol) was dissolved in DCM (55 mL). TFA (55 mL) and TIS (5.5 mL) were added and the solution was stirred for 2 hours in the dark at 25 °C. The DCM and TFA were evaporated under a light flow of nitrogen and the resulting residue was redissolved in EtOAc and washed with 10% $NaHCO_3$ (2 x 50 mL). The organic layer was dried over $MgSO_4$, filtered, and concentrated. The product was purified by silica gel flash chromatography (EtOAc, then 9:1 EtOAc/MeOH with 1% AcOH, $R_f = 0.18$) to give **10** (158 mg) in 64% yield. ^1H NMR (300 MHz, $CDCl_3$) δ ppm: 7.97-7.26 (m, 12H), 6.21 (bs, 1H), 6.13 (m, 1H), 5.62 (m, 1H), 5.01 (m, 2H), 4.51-4.11 (m, 5H), 2.70 (m, 2H), 1.75 (m, 3H), 1.30 (m, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ ppm: 170.1, 156.9, 147.1, 143.8, 141.5, 137.2, 134.2, 129.4, 128.0, 127.9, 127.3, 125.3, 124.8, 120.2, 116.6, 73.8, 67.6, 67.1, 63.0, 58.1, 47.3, 19.8, 18.5. ^{31}P NMR (121.5 MHz, $CDCl_3$) $\delta_f$ppm: -3.91. ESI-MS: $[MNa]^+$ 662.2857 (obsd), 662.2877 (calcd).

Quantum yield calculation. The quantum yield, Φ, of **cpChk2** was calculated by comparing the percent loss of the caged peptide after photolysis, as quantified by HPLC analysis, to that of a standard, caged phosphate (cp), with a known Φ.[17] HPLC analysis was performed prior to (t = 0 s) and immediately after (t = 15 s) irradiation. The analytical HPLC gradient used for the samples was 93:7 to 65:35 (water:MeCN with 0.1% TFA) over 15 minutes, then 65:35 to 40:60 over 20 minutes at 1.0 mL/minute. Samples were irradiated for 15 seconds at λ = 365 nm and an intensity of 7,330 μW/cm^2 on a DNA transilluminator in glass vessels with path lengths of 1 mm. Peptide solutions were prepared in water at pH 7.1 with 5 mM DTT and 0.46 mM inosine. The A_{350} value

of each sample was measured in water, pH 7.1. Below is the quantum yield calculation for **cpChk2**.

$$\Phi_{cpChk2} = \Phi cp \times (\%\Delta cpChk2 / \%\Delta cp) \times (A_{350}cp/A_{350}cpChk2)$$

$$\Phi_{cpChk2} = 0.54 \times (14.520 / 19.865) \times (0.02119 / 0.02516)$$

$$\mathbf{\Phi_{cpChk2} = 0.33}$$

Data table for ***cpChk2***

Peak Name	Retention Time (minutes)	Average Area	Avg Normalized Area	Avg A_{350} in water, pH 7.1
Inosine (t=0)	2.97	1.45	1.000	
cpChk2 (t=0)	21.33	3.34	2.294	0.02516
Inosine (t=15)	2.93	1.50	1.000	
cpChk2 (t=15)	21.28	2.95	1.9609	

Data table for caged phosphate

Peak Name	Retention Time (minutes)	Average Area	Avg Normalized Area	Avg A_{350} in water, pH 7.1
Inosine (t=0)	2.95	1.44	1.000	
Caged P (t=0)	18.37	1.91	1.334	0.02119
Inosine (t=15)	2.86	1.43	1.000	
Caged P (t=15)	18.35	1.72	1.069	

Acknowledgements I would like to acknowledge Debbie Rothman for spearheading the caged phosphopeptide project. Debbie established the syntheses for caged phosphoserine-containing peptides and Eugenio Vazquez developed the syntheses for caged phosphotyrosine-containing peptides.

References

1. Rothman, D. M., Vazquez, M. E., Vogel, E. M. & Imperiali, B. General Method for the Synthesis of Caged Phosphopeptides: Tools for the Exploration of Signal Transduction Pathways. *Org. Lett.* **4**, 2865-2868 (2002).

2. Rothman, D. M., Vazquez, M. E., Vogel, E. M. & Imperiali, B. Caged Phospho-Amino Acid Building Blocks for Solid-Phase Peptide Synthesis. *J. Org. Chem.* **68**, 6795-6798 (2003).
3. Walsh, C. T. *Posttranslational modification of proteins: expanding nature's inventory* (Roberts and Company, Englewood, 2006).
4. Lauffenburger, D. A. & Horwitz, A. F. Cell migration: a physically integrated molecular process. *Cell (Cambridge, Mass.)* **84**, 359-69 (1996).
5. Rothman, D. M., Shults, M. D. & Imperiali, B. Chemical approaches for investigating phosphorylation in signal transduction networks. *Trends Cell Biol.* **15**, 502-510 (2005).
6. Gurney, A. M. & Lester, H. A. Light-flash physiology with synthetic photosensitive compounds. *Physiol. Rev.* **67**, 583-617 (1987).
7. Kaplan, J. H. & Ellis-Davies, G. C. R. Photolabile chelators for the rapid photorelease of divalent cations. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6571-5 (1988).
8. Walker, J. W. et al. Signaling pathways underlying eosinophil cell motility revealed by using caged peptides. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1568-1573 (1998).
9. Wood, J. S., Koszelak, M., Liu, J. & Lawrence, D. S. A Caged Protein Kinase Inhibitor. *J. Am. Chem. Soc.* **120**, 7145-7146 (1998).
10. Walker, J. W., Reid, G. P., McCray, J. A. & Trentham, D. R. Photolabile 1-(2-nitrophenyl)ethyl phosphate esters of adenine nucleotide analogs. Synthesis and mechanism of photolysis. *J. Am. Chem. Soc.* **110**, 7170-7 (1988).
11. Kaplan, J. H., Forbush, B., 3rd & Hoffman, J. F. Rapid photolytic release of adenosine 5'-triphosphate from a protected analogue: utilization by the Na:K pump of human red blood cell ghosts. *Biochemistry* **17**, 1929-35 (1978).
12. Haugland, R. P. *Handbook of Fluorescent Probes and Research Products* (ed. Gregory, J.) (Molecular Probes, Eugene, OR, 2005).
13. Kupihar, Z., Varadi, G., Monostori, E. & Toth, G. K. Preparation of an asymmetrically protected phosphoramidite and its application in solid-phase synthesis of phosphopeptides. *Tetrahedron Lett.* **41**, 4457-4461 (2000).

14. Meutermans, W. D. F. & Alewood, P. F. A simple and effective procedure for the synthesis of the 'difficult' phosphotyrosine-containing peptide Stat 91 (695-708). *Tetrahedron Lett.* **37**, 4765-4766 (1996).

15. Yaffe, M. B. & Smerdon, S. J. Phosphoserine/threonine binding domains: you can't pSERious? *Structure (Cambridge, MA, United States)* **9**, R33-R38 (2001).

16. Yaffe, M. B. & Elia, A. E. H. Phosphoserine/threonine-binding domains. *Curr. Opin. Cell Biol.* **13**, 131-138 (2001).

17. Ellis-Davies, G. C. R. & Kaplan, J. H. Nitrophenyl-EGTA, a photolabile chelator that selectively binds Ca2+ with high affinity and releases it rapidly upon photolysis. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 187-91 (1994).

18. Humphrey, D. et al. In Situ Photoactivation of a Caged Phosphotyrosine Peptide Derived from Focal Adhesion Kinase Temporarily Halts Lamellar Extension of Single Migrating Tumor Cells. *J. Biol. Chem.* **280**, 22091-22101 (2005).

19. Nguyen, A., Rothman, D. M., Stehn, J., Imperiali, B. & Yaffe, M. B. Caged phosphopeptides reveal a temporal role for 14-3-3 in G1 arrest and S-phase checkpoint function. *Nat. Biotechnol.* **22**, 993-1000 (2004).

20. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411-23 (1995).

Chapter 3

Semisynthesis of caged phosphoTyr31 paxillin

Introduction

N^{α}-Fmoc-protected 1-(2-nitrophenyl) ethyl caged phosphoserine, phosphothreonine, and phosphotyrosine building blocks (Chapter 2) facilitate the straightforward assembly of any caged phosphopeptide through Fmoc-based SPPS.[1] Caged phosphopeptides are valuable tools for biological studies in which the phosphorylated species is active at the peptide level, and these probes have been utilized successfully for both in vitro and cellular studies.[2] A powerful extension of the caged phosphopeptide methodology would be the incorporation of caged phosphorylated residues into complex, multi-domain proteins. The resulting full-length probes would allow researchers to dissect of the role of individual phosphorylation sites in "real time", while preserving other phosphorylation sites, binding and localization determinants, or enzymatically-active domains within the target proteins.

A full-length protein target of significant interest to our lab is a caged phosphorylated analog of paxillin, a 61-kDa cytoplasmic phosphoprotein. Paxillin localizes to focal adhesions, which are sites of cellular contact with the extracellular matrix, and is implicated in cellular adhesion and cell migration (**Fig. 3.1**).[3] The control of cellular adhesion and migration is central in the regulation of biological processes, including embryogenesis, wound repair, and metastasis.[4] Paxillin is a multi-domain protein that contributes to the control of these processes by acting as a dynamic scaffold for signaling and structural proteins. Phosphorylation of paxillin at specific serine, threonine or tyrosine residues spanning the molecule creates distinct protein binding sites and thereby directs paxillin localization to focal adhesions, or influences the controlled

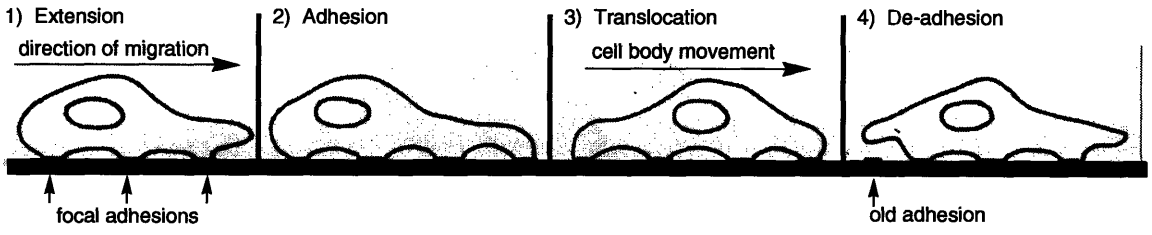


Figure 3.1. Cellular migration.[5] The movement of cells involves the tightly regulated formation and subsequent disassembly of focal adhesions.

assembly or dissolution of signaling cascades.[6] Specifically, phosphorylation of paxillin at Tyr31 creates a binding platform for a Src homology 2 (SH2) domain, facilitating interaction with SH2 domain-containing proteins such as Crk, an adaptor protein that recruits other signaling molecules involved in cellular migration.[7]

In this chapter we describe the semisynthesis of a paxillin analog with a caged phosphotyrosine (cpTyr) at position 31 of the 557-residue protein, which will be used as a tool to investigate the impact of Tyr31 paxillin phosphorylation on cellular migration (**Fig. 3.2**). The probe was constructed using NCL to install a caged phosphopeptide on the *N*-terminus of a biologically expressed fragment of paxillin. To serve as non-phosphorylated and discreetly phosphorylated biological controls, paxillin variants with a Tyr or phosphoTyr (pTyr) residue at position 31 were constructed analogously to the caged phosphorylated construct.

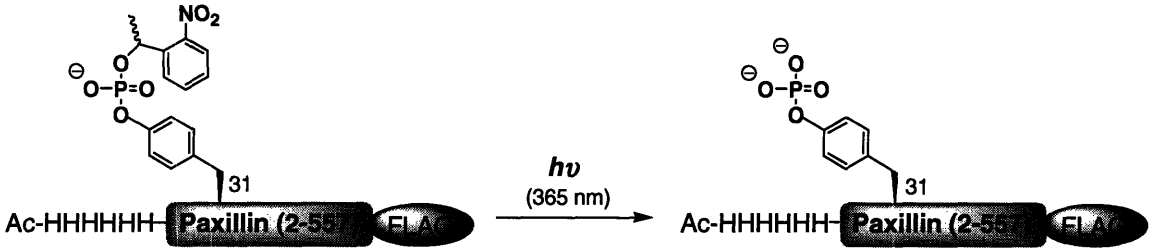


Figure 3.2. Caged phosphoTyr31 paxillin. Uncaging with long-wavelength UV light releases phosphoTyr31 paxillin. The proteins are shown with *N*-terminal hexahistidine and *C*-terminal FLAG tags

Paxillin structure and overview of selected interactions

Structurally, paxillin includes numerous binding domains that facilitate a range of protein-protein interactions (**Fig. 3.3**). Near the *N*-terminus of paxillin, there are five LD motifs, which are 8-residue peptide sequences that begin with a leucine-aspartic acid (L-D) pairing. Modeled as an α-helix, leucine or valine residues lie along one side of these motifs, presenting a hydrophobic face.[8] The LD repeats bind a common domain, the PBS (paxillin binding subdomain), conserved in various binding partners, including the actin-binding proteins actopaxin and vinculin, focal adhesion kinase (FAK), a tyrosine kinase

that phosphorylates paxillin, and paxillin kinase linker (PKL), a protein involved in actin cytoskeleton dynamics.[3] Also on the *N*-terminal half of paxillin, several proline-rich regions provide binding sites for SH3 domains found in members of the Src family. SH3 domains are modular adaptor domains that bind with μM affinity to proline-rich segments on the surface of target proteins. Additionally, phosphorylation of paxillin on Tyr31 and Tyr118 by Src and FAK creates two binding sites for SH2 domains, enabling interaction with the adaptor protein Crk, as mentioned above, and Csk (C-terminal Src-kinase), a tyrosine kinase that negatively regulates Src activity. SH2 domains are phosphorylation-sensitive adaptor domains that bind peptide sequences with a pTyr-X-X-Φ consensus motif, where X is any amino acid except proline, and Φ is a hydrophobic residue. The extent of paxillin phosphorylation is regulated by interaction with the extracellular matrix (ECM) and extracellular environmental signals provided by growth factors. Other sites of phosphorylation include Ser188 and Ser190, which are phosphorylated subsequent to cellular adhesion to fibronectin.[6]

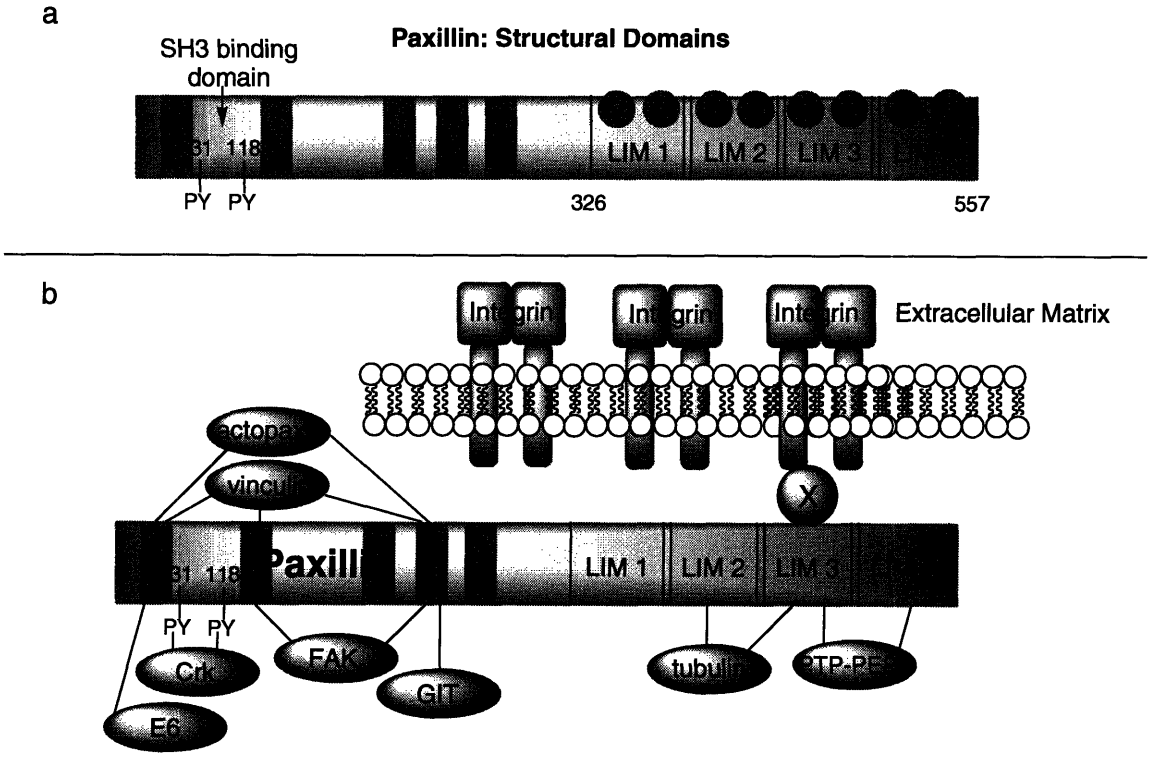


Figure 3.3.[6,9] **a**) Paxillin structural domains. The *N*-terminus includes LD repeats and Pro-rich SH3 binding regions. The *C*-terminus comprises four LIM domains. SH2-binding regions are created by Tyr31 and Try118 phosphorylation. **b**) Selected paxillin-interacting proteins, including signaling proteins and structural actin-binding proteins.

The *C*-terminus of paxillin includes four LIM domains, cysteine/histidine-rich double zinc fingers that serve as protein-protein binding sites. These domains interact with both kinases and phosphatases. One such LIM-binding protein is PTP-PEST,[10] a phosphatase that triggers focal adhesion disassembly by dephosphorylating CAS (Crk-associated substrate), a protein recruited to the area through its interaction with Crk. The LIM domains, specifically LIM3, also function as focal-adhesion targeting sites in paxillin, potentiated by serine/threonine phosphorylation.[11]

As suggested by the structural and regulatory roles of the numerous binding partners described, paxillin is instrumental in the regulation of cell motility, both by direct interaction with structural elements, as well as by functioning as a scaffold to bring various signaling proteins into close proximity, facilitating interaction. Paxillin phosphorylation is critical for the regulation of a number of these processes. For example, phosphorylation of paxillin at Tyr31 and Tyr118 has been shown to increase during cell spreading.[12] In a related study, a Tyr31Phe/Tyr118Phe mutant was used to verify the necessity of phosphorylation at those sites for EphB1 receptor mediated cell migration.[13] Additionally, signaling through these residues has recently been shown to be involved in the PTP-PEST-dependent regulation of cellular mobility.[14]

The involvement of paxillin in cytoskeletal dynamics and cellular migration may be of great consequence in cancer metastasis. The E6 oncoprotein from papillomavirus, which leads to cervical carcinomas, binds to the LD motifs of paxillin. This decreases interaction with normal LD binding partners, including vinculin and FAK, likely contributing to the disruption of the actin cytoskeleton, which occurs in papillomavirus-infected cells.[15] Additionally, oncongenic equivalents of Src (v-Src) and Crk (v-Crk) have been identified, which similarly perturb normal paxillin binding. Specifically, these mutants disrupt the regulation of the MAPK (mitogen-activated protein kinase) cascade involved in controlled cell proliferation and gene expression.[16] Phosphorylation of paxillin has also been directly linked to tumor metastasis. In an investigation of the invasive capacity of rat ascites hepatoma MM1 cancer cells, it was found that paxillin phosphorylation on Tyr31 and Tyr118 was required for transcellular migration.[17]

Both for the involvement of paxillin modification in metastasis and in the regulation of normal cell processes, there is enormous interest in the downstream cellular

effects of paxillin phosphorylation. Specifically, Tyr31 phosphorylation is involved in a number of regulatory processes, but little is known about the effect of phosphorylation in "real time," and little work had been done to dissect the isolated effect of Tyr31 phosphorylation from the compound effect of Tyr31/Tyr118 phosphorylation. A photolabile precursor of phosphoTyr31 was selected as the target for semisynthesis to probe the role of Tyr31 on cell migration. The proximity of a critical phosphorylated residue to the *N*-terminus made it a promising candidate for semisynthesis through a single NCL reaction.

Results and Discussion

3-1. Overview of Paxillin Semisynthesis

The semisynthesis of paxillin (**Y31Pax**), phosphoTyr31 paxillin (**pY31Pax**), and caged phosphoTyr31 paxillin (**cpY31Pax**) was accomplished by the ligations of synthetic thioesters corresponding to residues 2-36 of paxillin to a biologically expressed segment comprising residues 37-557 (**Fig 3.4**). An Asn37Cys mutation was designed to provide a Gly36-Cys37 junction for NCL. This site was chosen because it is in a region of paxillin free of any predicted secondary structure, and because the presence of a glycine as the terminal thioester residue is known to increase ligation efficiency.[18] Three peptide thioesters corresponding to residues 2-36 of paxillin were synthesized, with a Tyr, pTyr, or cpTyr building block at residue 31 (**Fig. 3.4a, 18**). An *N*-terminal hexahistidine tag was included in the peptides to provide a handle for visualization, as well as purification of the ligation product away from any unreacted starting material following ligation.

For the *C*-terminal fragment of paxillin, residues 38-557 were expressed with an *N*-terminal purification tag and a protease cleavage sequence (PCS) positioned immediately preceding the paxillin insert. Treatment of the precursor protein with a protease resulted in cleavage of the purification tag and exposure of an *N*-terminal cysteine residue to produce Cys-paxillin(38-557) for subsequent ligation. Ligation of the *C*-terminal fragment with the synthesized thioesters **18a**, **18b**, and **18c** yielded **Y31Pax, pY31Pax**, and **cpY31Pax** respectively.

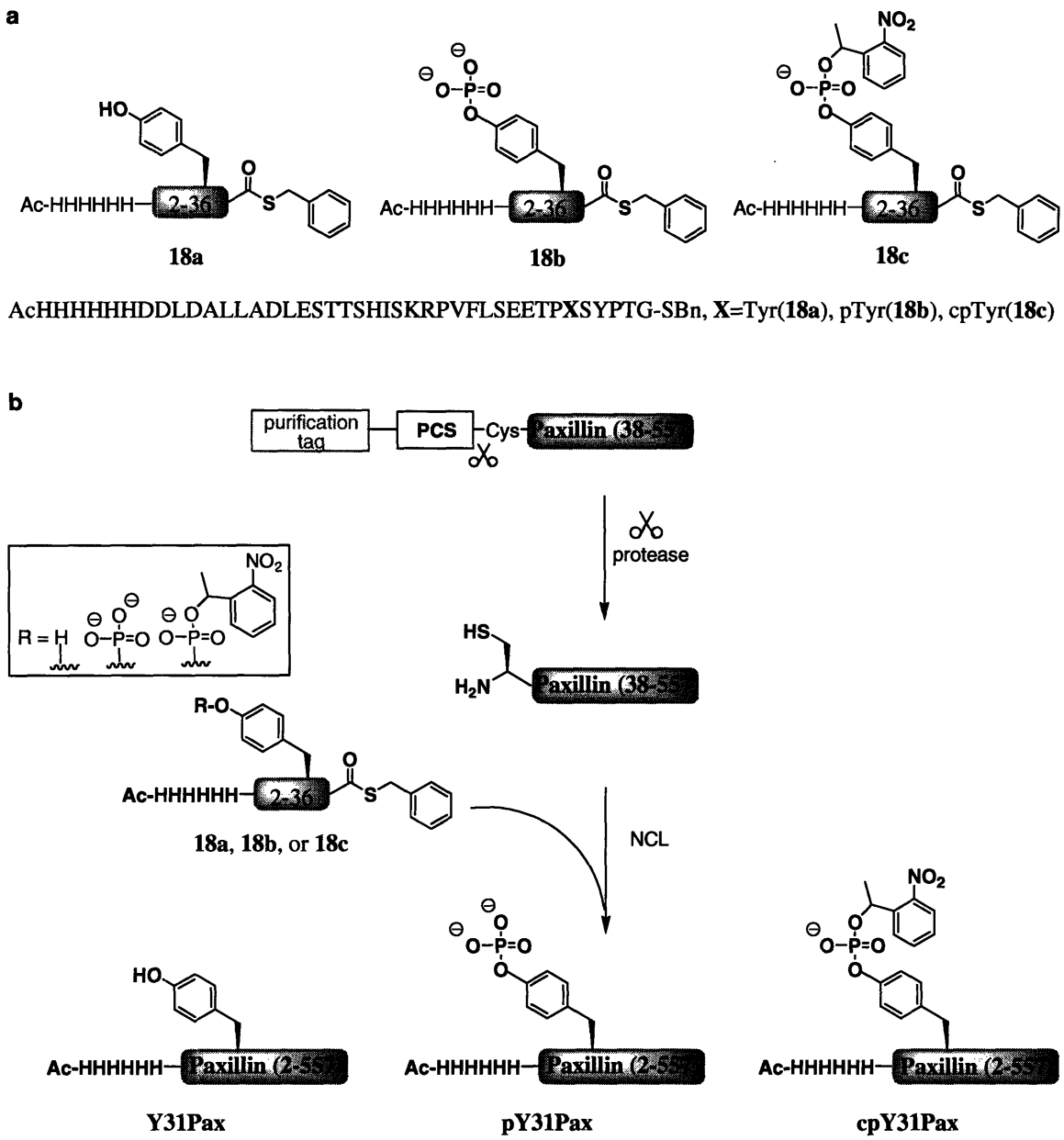


Figure 3.4. Semisynthesis of paxillin analogues. (**a**) Synthetic peptide thioesters corresponding to the *N*-terminus of paxillin. (**b**) Semisynthetic strategy. PCS = protease cleavage site.

3-2. Synthesis of paxillin peptide thioesters

For the synthesis of the 41-residue *N*-terminal thioesters (Ac-HHHHHH-DDLDALLADLESTTSHISKRPVFLSEETP-X-SYPTG, X = Tyr (**18a**), pTyr (**18b**), or cpTyr (**18c**)), the peptides were prepared on highly acid labile TGT resin using Fmoc-

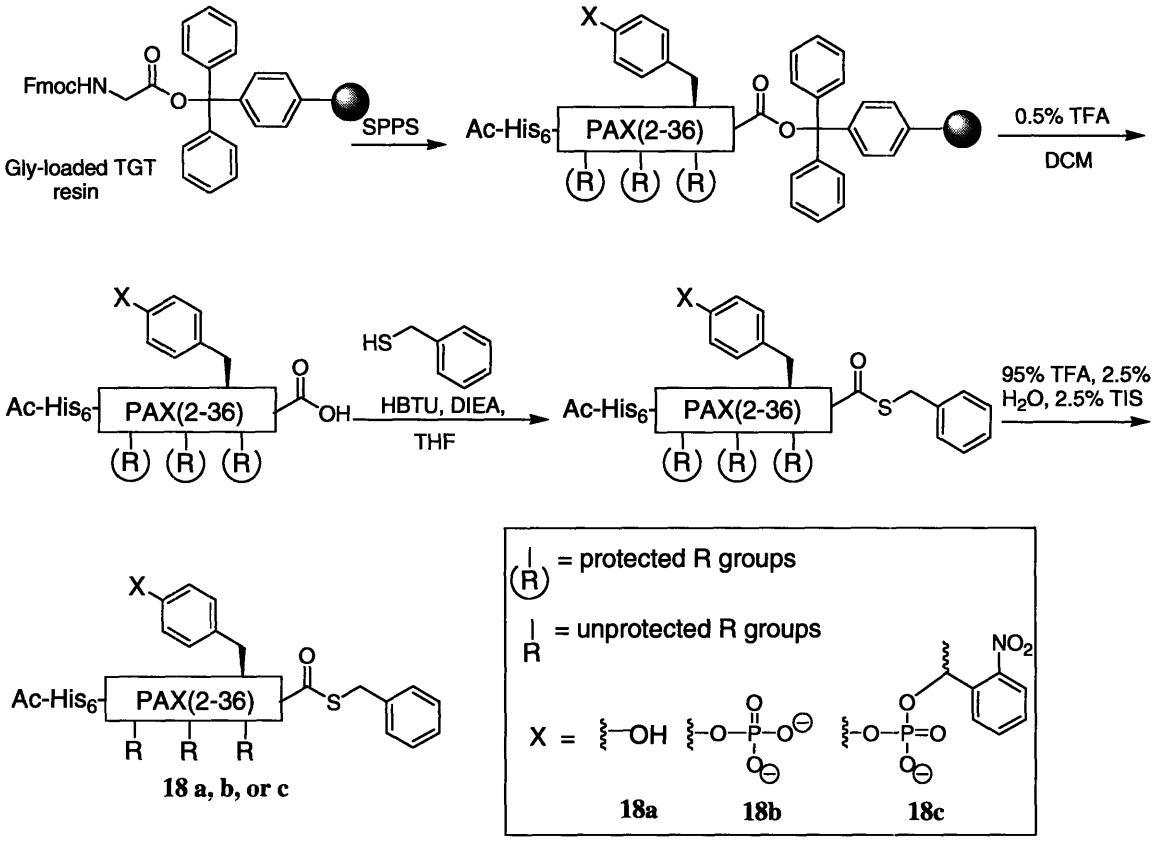
based SPPS. The cpTyr building block, synthesized as described in Chapter 2, was used to install a pTyr masked by the 1-(2-nitrophenyl)ethyl (NPE) caging group in peptide **18c**. In a procedure modified from published methods,[19,20] the peptides were cleaved from the resin with 0.5% TFA to release fully side-chain protected peptides as *C*-terminal carboxylic acids. The *C*-termini were activated with HBTU and DIPEA and reacted with benzylmercaptan to yield the corresponding thioesters. The amino acid side chains were subsequently deprotected with TFA in the presence of scavengers to produce **18a**, **18b**, and **18c** (**Scheme 3.1**). The thioesters were purified by reverse phase HPLC and stored as lyophilized powders at -20 °C for subsequent use in NCL. Between 10 and 30 mg of each thioester variant were synthesized.

Prior to the large-scale thioester syntheses, an initial test synthesis of residues 1-36 of paxillin was carried out to identify problematic coupling steps. It was important to maximize the efficiency of the peptide synthesis, since the desired thioesters are long and require "off-bead" thioesterification following the SPPS. The test peptide, Pax(1-36) was constructed on an automated peptide synthesizer with an acetic anhydride capping step following each coupling. Samples were removed from the synthesizer subsequent to the coupling of Leu25, Lys20, Glu12, and Met1, resulting in a 12-, 17-, 25-, and 36-residue peptide respectively. Test cleavages of the peptide fragments were monitored by analytical HPLC followed by ESMS. Analysis of the 12 and 17-residue peptides revealed a single major peak by HPLC with a mass corresponding to the desired peptide. However, the 25 and 36-residue fragments produced more complicated spectra, each with a second large HPLC peak (m/z 1107.5) in addition to the peak corresponding to the expected product. Based on the mass, it was determined that the probable identity of the extra peak was Ac-ISKRPVFLSEETPYSYPTG-COOH, resulting from an unsuccessful coupling of Fmoc-His(Trt)-OH to Ile18, leading to capping of the19-residue peptide. It is reasonable that this would be a difficult coupling step due to steric interactions between the bulky trityl protecting group on histidine and the β-branched isoleucine residue.

In subsequent syntheses of the paxillin thioester variants **18a**, **b** and **c**, the challenging histidine/isoleucine coupling was performed manually using benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP) and DIPEA as activating agents. The absence of free amines was confirmed by a TNBS test and the

peptides were completed on a peptide synthesizer to give NH_2-Pax(2-36). The hexahistidine tag and acetyl cap were added manually to complete the peptides, which were subsequently converted to thioesters as previously described.

Scheme 3.1. Synthesis of paxillin thioesters.



3-3. Expression of the C-terminal precursor for NCL

There are several strategies for accessing biochemically-expressed protein fragments with an *N*-terminal cysteine residue for use in NCL (see Chapter 1 for a complete discussion). These approaches include intein-based strategies and exogenous protease cleavage of a precursor protein. A major disadvantage to an intein method is the common complication of spontaneous intein cleavage, while a minor drawback is the need for a specialized vector. A more straightforward method, compatible with virtually any vector, is the use of a precursor protein expressed with a cysteine residue

immediately *C*-terminal to a protease cleavage site. In vitro cleavage of the appropriate precursor protein generates the requisite *N*-Cys protein fragment.

For the generation of the *C*-terminal paxillin fragment, Cys-Pax(38-557), a hexahistidine-fusion protein of Pax(38-557) with a protease cleavage site and a cysteine flanking the paxillin insert was designed. The DNA encoding His_6-PCS-Cys-Pax(38-557) (PCS = protease cleavage site, discussed in detail in Section 3-4) was transformed into BL21(DE3) *E. coli* cells, expressed, and purified on a Ni/NTA column.

Paxillin is known to be a challenging protein to express, with poor yields and high levels of degradation. Consistent with such reports, attempts at expressing the paxillin construct resulted in barely detectable levels of paxillin expression, visible only by sensitive Western blot analysis and accompanied by a large number of degradation and/or truncation products (**Fig. 3.5**). Ultimately, to access sufficient quantities of protein and overcome truncation difficulties, the protein was fermented in 10-L batches using codon-enhanced cells and the construct was reengineered to incorporate both *C*-terminal and *N*-terminal purification tags. These two key changes, which dramatically increased yield and purity of the paxillin construct, are discussed below.

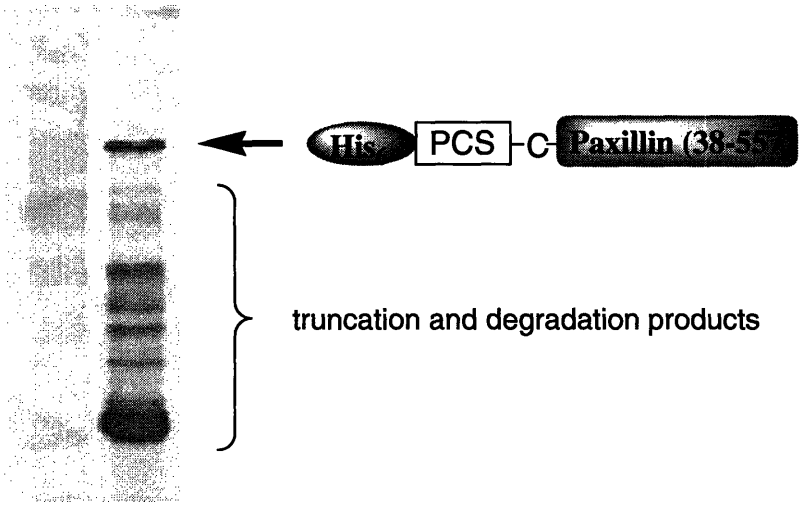


Figure 3.5. Representative paxillin expression in the BL21 cell line. Western blot, probed with an anti-paxillin primary antibody, of Ni/NTA purified His_6-PCS-Cys-Pax(38-557). The full construct is indicated by an arrow.

Use of a codon-enhanced E. coli *cell line*

An examination of the paxillin(38-557) DNA revealed the presence of a significant number of rare codons. The insert includes 27 rare proline, 1 rare isoleucine, 2 rare arginine, and 6 rare glycine codons, totaling 7% of the insert DNA. Further, paxillin includes several proline-rich regions, meaning those codons tend to be clustered together, exacerbating the strain on the translational machinery. This suggests that many of the originally suspected degradation products were actually truncation products due to excessive demand for tRNAs that are in low abundance in *E. coli* cells.

To circumvent the problem of rare codons, the paxillin DNA was transformed into codon-enhanced cells, which contain plasmids that encode rare *E. coli* tRNAs. Initially the Novagen Rosetta (DE3) cells were used, which express all of the rare *E. coli* tRNAs on a plasmid with chloramphenicol resistance. Unfortunately, standard glycerol preparations of these cells failed to inoculate subsequent expressions. There have been reports of plasmid instability for inexplicable reasons with Rosetta cells from greater than 10% glycerol stocks. In our experience there were plasmid instability issues despite minimal glycerol concentrations. Since the Rosetta system was the only option offering more than three rare codons in a single codon-enhanced cell line, it is possible that the observed instability was a result of problematically large plasmids.

The paxillin vector was next transformed into the Stratagene BL21-CodonPlus®(DE3)-RP competent cells, which code for the AGG (Arg), AGA (Arg), and CCC (Pro) tRNAs. Expression of the paxillin construct was significantly improved over that achieved with BL21(DE3) cells, with cleaner gels and higher yields. However, there was still a major truncation product at about 25 kDa, and there were several other undesired bands visible by Coomassie and Western blot analysis. A number of parameters were optimized, including the media (LB or TB), temperature for expression (16, 24, or 37 °C), OD before induction (0.4 to 0.9), IPTG concentration (0.001, 0.01, 0.1 or 1 mM), and expression time (3, 4, 5, 7, or 20 hours). Induction with 0.1 mM IPTG at an OD of 0.6 for 24 hours at 16 °C was identified as the optimum set of conditions for growth, although these conditions do not completely prevent the 25-kDa truncation product or result in high yields of protein. Ultimately it was determined that the most significant hurdle remaining was not the modest yield or the truncation product, since *E.*

coli expression is easy and inexpensive to scale up, but the lack of a second purification handle that would allow clean separation of the desired protein from the truncation and degradation products.

Construct redesign: incorporation of a second purification tag

A GST-fusion protein of Pax(38-557) with a *C*-terminal FLAG tag was designed to increase protein production and enable purification via the GST and FLAG tags (**Fig. 3.6**). Since protein biosynthesis proceeds from the amino to the carboxyl terminus, none of the truncation products should contain the FLAG moiety. The FLAG-affinity resin,

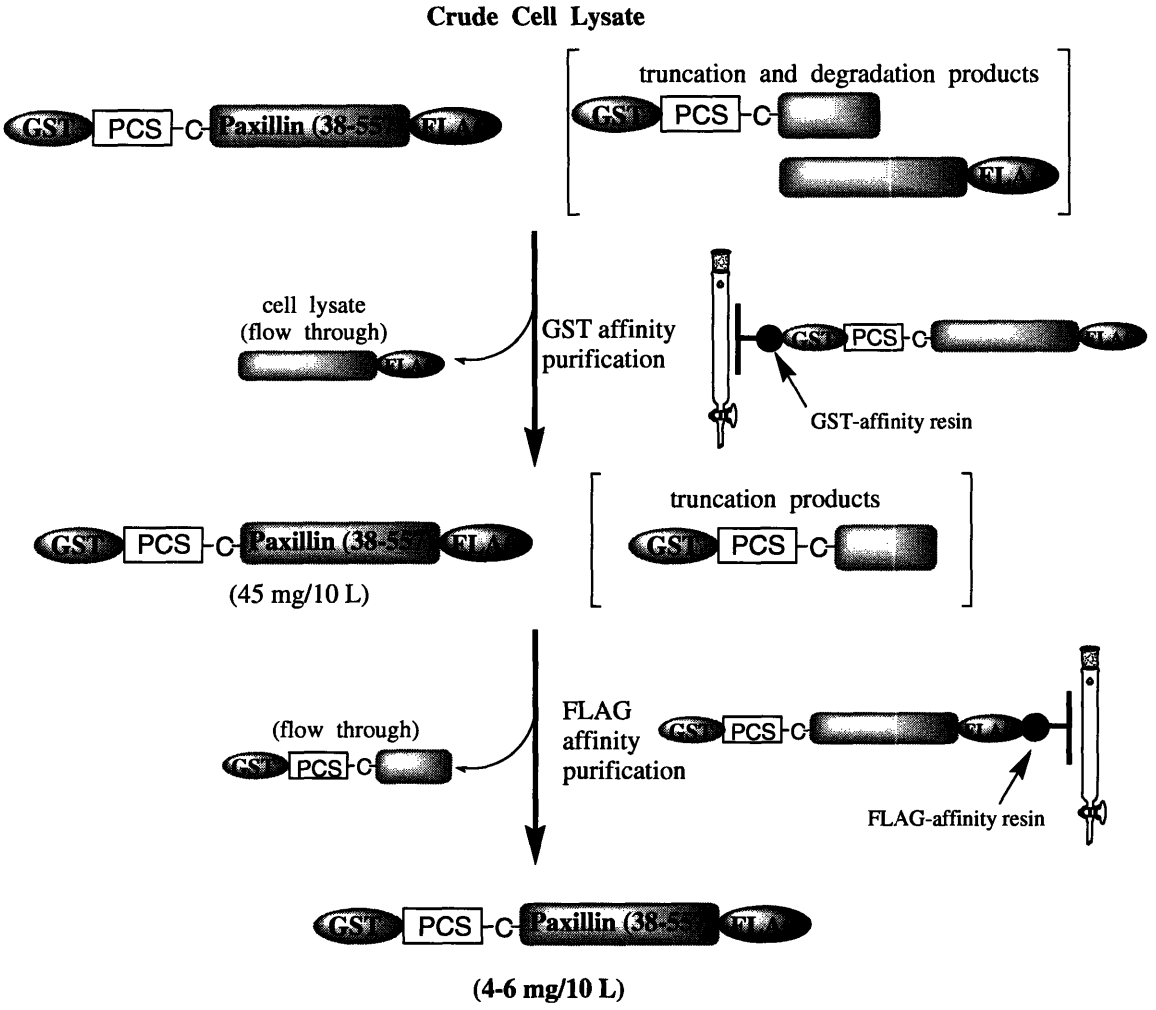


Figure 3.6. Complete isolation of GST-PCS-Cys-Pax(38-557)-FLAG from the truncation and degradation products is possible by affinity purification via both the *N*-terminal and *C*-terminal tags.

which is several times more expensive than the GST-affinity resin, retains binding capacity significantly longer when the protein loaded is partially purified, so the *N*-terminal GST-affinity purification was performed first.

A Pax(38-557) insert was prepared for cloning into a pGEX-4T-2 GST-fusion vector using primers to encode a FLAG tag at the *C*-terminus and a protease cleavage site immediately N-terminal to the expressed PAX(38-557) fragment. After PCR and cloning, the isolated plasmid DNA from an appropriate miniprep was transformed into BL21-CodonPlus®(DE3)-RP competent cells for the expression of GST-PCS-Cys-Pax(38-557)-FLAG. Protein expression was carried out in LB-carbenicillin-chloramphenicol with 0.1 mM IPTG induction, followed by 24 hours of fermentation at 16 °C. After initial test expressions and purifications, subsequent expressions were carried out by fermentation in 10-L batches. Gel analysis of the crude expression showed a major truncation product, as seen with the original paxillin construct.

Crude GST-PCS-Cys-Pax(38-557)-FLAG was purified on GST and FLAG affinity columns to achieve complete separation from the truncation product (**Fig. 3.7**). The crude lysate was first purified by batch purification using a GST column and eluted from the column using 50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0. A 10-L

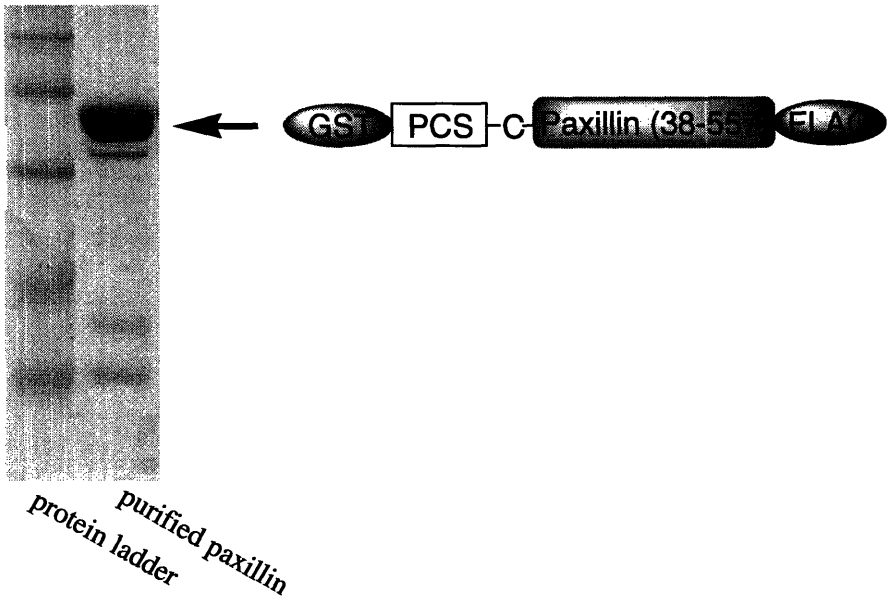


Figure 3.7. Purified GST-paxillin fusion protein following expression in codon-enhanced cells and purification via the *N* and *C*-terminal tags.

fermentation of GST-PCS-Cys-Pax(38-557)-FLAG resulted in 45 mg of protein after the GST-affinity purification, as quantified using a BioRad protein assay, which is compatible with the presence of glutathione. Prior to purification on an anti-FLAG column, the GST-purified protein was dialyzed into TBS to remove the glutathione used for elution. The FLAG affinity resin is not compatible with reducing agents, such as glutathione, which could reduce the disulfide bonds in the M2 antibody on the resin. The dialyzed protein solution was purified on the FLAG column, again using a batch method for binding to the resin. The FLAG fusion protein was eluted from the column with 1.2-mL aliquots of 0.1 M glycine HCl, pH 3.5 and collected into eppendorf tubes containing 160 μL of Tris-HCl, pH 8.0 and 100 μL of 10 X TBS. The yield of GST and FLAG-purified protein was 4 to 6 mg from a 10-L fermentation. Since expressions in *E. coli* are economical to scale up, obtaining high quantities of pure product was straightforward.

3-4. Protease selection

There are several commercially available proteases that cleave *C*-terminal to their recognition sequences, such as Factor Xa,[21] which has previously been used to generate *N*-terminal cysteine-containing proteins for NCL, as well as enterokinase, ubiquitin *C*-terminal hydrolase, and furin, which should have similar utility. Other widely used proteases, including thrombin, cannot be used in NCL applications because they require recognition elements *C*-terminal to their cleavage site, precluding the ability to reveal an *N*-terminal cysteine residue. TEV protease, a highly selective cysteine protease, typically recognizes a serine or a glycine in the P1' site, but also will accept a cysteine residue at that position, making it another option for the generation of *N*-terminal cysteine containing fragments.[22]

The original construct for the paxillin *C*-terminal fragment precursor was designed with a Factor Xa recognition site. Factor Xa was initially selected based on its prevalence in NCL applications as well as its use within the Imperiali group for the removal of affinity tags. However, despite a range of cleavage conditions tested, Factor Xa treatment consistently resulted in unselective proteolysis of paxillin. Below are described the paxillin test reactions with Factor Xa and the subsequent screening of other

proteases that lead to the selection of TEV protease for cleavage of GST-PCS-Cys-Pax(38-557)-FLAG.

A range of cleavage conditions were tested on the GST paxillin fusion GST-IEGR-Cys-Pax(38-557)-FLAG, which included the Factor Xa recognition motif, IEGR. Upon successful cleavage of the paxillin construct, gel analysis should have revealed a loss of the band corresponding to the GST-paxillin fusion (83 kDa) and the appearance of two new bands from Cys-Pax(38-557)-FLAG (56 kDa) (**19**) and the liberated GST tag (27 kDa).

The manufacturer reports 0.2 μg of Factor Xa proteolyze 95% of 10 μg of target protein at room temperature in 6 hours. Initially, 50 to 100-μL test reactions were carried out at room temperature on 10 μg of purified protein with 0.25, 0.5, 1, and 3 μg of protease. In all reactions, Western blot analysis revealed the loss of GST-IEGR-Cys-Pax(38-557)-FLAG within three hours, resulting exclusively in species with apparent masses well below the expected 56 kDa (**Fig. 3.8a**). Further test cleavages were carried out with increasingly smaller amounts of protease, 0.2 to 0.0125 μg per 10 μg of protein, in an attempt to prevent proteolysis at secondary sites within paxillin. Aliquots were analyzed for each concentration of protease after 1, 3, 6, and 20 hours with the goal of establishing conditions to achieve sufficient cleavage at the primary protease site without significant cleavage at secondary sites. As the amount of Factor Xa was incrementally scaled down, the 83-kDa-protein (GST-IEGR-Cys-Pax(38-557)-FLAG) disappeared more slowly. However, no appropriate band corresponding to the 56-kDa fragment could be detected at any of the time points. Comparable reactions were carried out at 4 °C, and again no band above 45 kDa was visible. Controls in which protein was incubated in the reaction conditions in the absence of Factor Xa confirmed that degradation of the paxillin fragment was a direct result only of the added protease.

As a final attempt, it was reasoned that secondary site cleavage might be reduced by the addition of zinc, which could improve folding of the paxillin fragment by organizing the four paxillin LIM domains, double zinc fingers that bind with nM affinity to a total of 8 zinc ions. It is possible that the LIM domains were stripped of their zinc in the purification process. Proteolysis reactions were run at room temperature and at 16 °C at several Factor Xa concentrations with either 200 or 600 μg ZnCl (50 to 150

equivalents.) Disappointingly, Western blot analysis of the reactions failed to reveal the desired product **19**.

While paxillin does not contain other IEGR repeats, a glycine-arginine dipeptide appears in several places in the sequence. Unwanted proteolysis by Factor Xa has been reported to occasionally occur after glycine-arginine pairs or other basic residues in a target protein.[23] Although this is most common following Gly-Arg-X sites, where X = isoleucine or threonine, which are not present in paxillin, it is possible that the glycine-arginine dipeptides are in highly exposed regions of the protein, making them more susceptible to cleavage. It was therefore determined that the paxillin construct must be

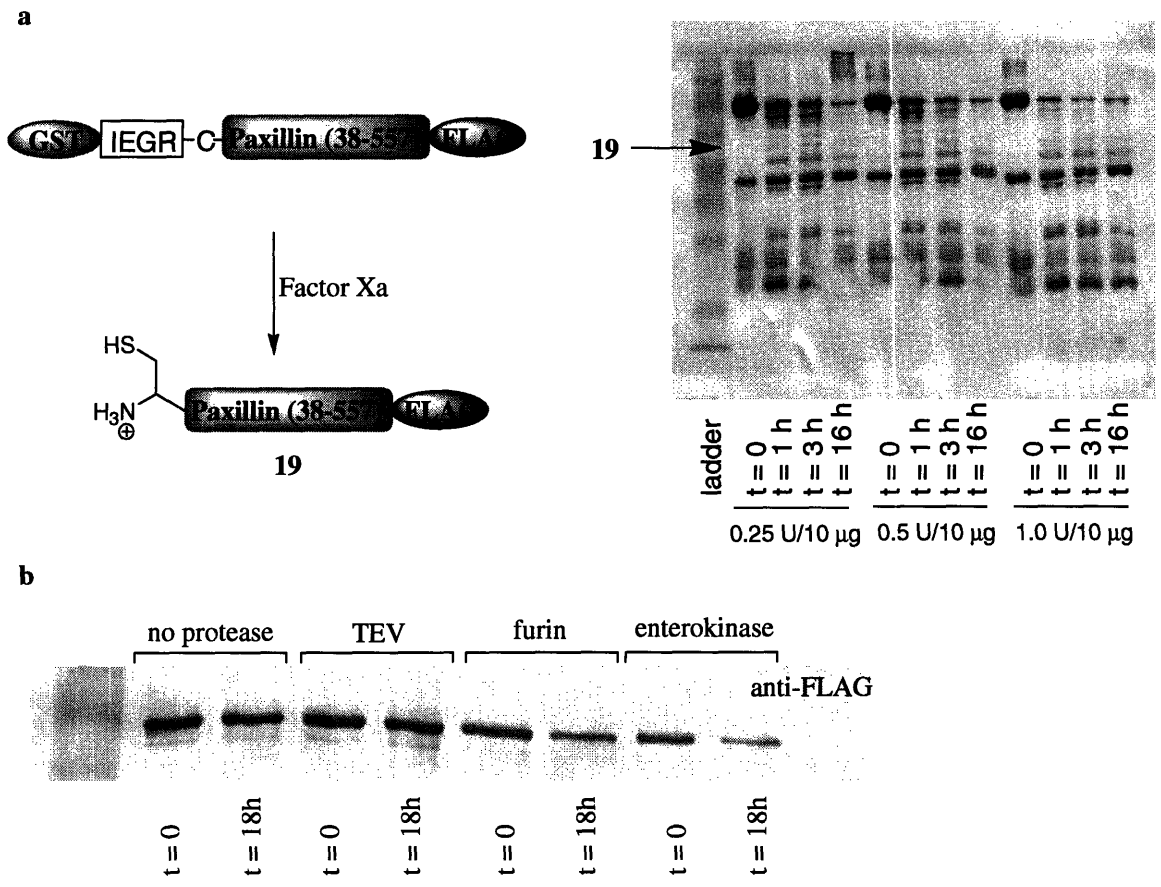


Figure 3.8. **a**) Treatment of GST-IEGR-Cys-Pax(38-557)-FLAG with Factor Xa. The Western blot to the right was probed with an anti-FLAG primary antibody and shows Factor Xa cleavage with varying amounts of protease monitored at time = 0, 1, 3, and 16 hours. The arrow indicates the expected location of the cleavage product, **19**. **b**) Screening for nonspecific paxillin degradation by proteases. An anti-FLAG-probed Western blot monitoring the treatment of GST-IEGR-Cys-Pax(38-557)-FLAG with enterokinase, furin, TEV, and a negative control with no protease.

redesigned to incorporate an alternative protease cleavage site. Proteases with orthogonal recognition determinants were therefore tested for paxillin compatibility.

In order to screen for a compatible protease, GST-IEGR-Cys-Pax(38-557)-FLAG was incubated with a number of proteases to identify, prior to cloning in a new cleavage site, proteases that would not cleave elsewhere in the paxillin fragment (**Fig. 3.8b**). In each case, the purified paxillin derivative was treated with the protease in the most concentrated conditions recommended by the respective manufacturer. The first protease tested was enterokinase, which cleaves *C*-terminal to a DDDDK recognition sequence. The *C*-terminal FLAG tag of GST-IEGR-Cys-Pax(38-557)-FLAG includes the DDDDK motif, but these are the terminal 5 residues, so enterokinase treatment should have no effect on the protein. Unfortunately significant proteolysis of the paxillin fragment was observed in the presence of enterokinase. Similarly, treatment of GST-IEGR-Cys-Pax(38-557)-FLAG with furin, which cleaves following RX(K/R)R, resulted in degradation of the protein, although to a lesser extent than with enterokinase or Factor Xa. Finally TEV protease, which cleaves *N*-terminal to the cysteine in its ENLYFQ(✂)C cleavage sequence, was tested and found to have no undesired cleavage with GST-IEGR-Cys-Pax(38-557)-FLAG.

Therefore a new construct was designed to incorporate a TEV cleavage site *N*-terminal to the paxillin insert. GST-ENLYFQ-Cys-Pax(38-557)-FLAG (**20**) was expressed and purified as described in Section 3-3. For TEV cleavage of **20**, the purified protein was diluted to 0.5 to 1 mg/mL into a TEV cleavage buffer with a final concentration of 50 mM Tris pH 8.0, 500 μM EDTA, and 5 mM beta-mercaptoethanol (BME). Standard TEV cleavage conditions call for dithiothreitol (DTT) in the cleavage mixture, but DTT can promote unwanted thioester hydrolysis in the subsequent ligation, so BME was used in its place. Nearly complete cleavage of the GST tag, revealing an *N*-terminal cysteine residue, was accomplished by treatment with TEV protease to yield Cys-Pax(38-557)-FLAG (**19**) (**Fig. 3.9**). The protein was concentrated to 2 to 4 mg/mL and used immediately in NCL.

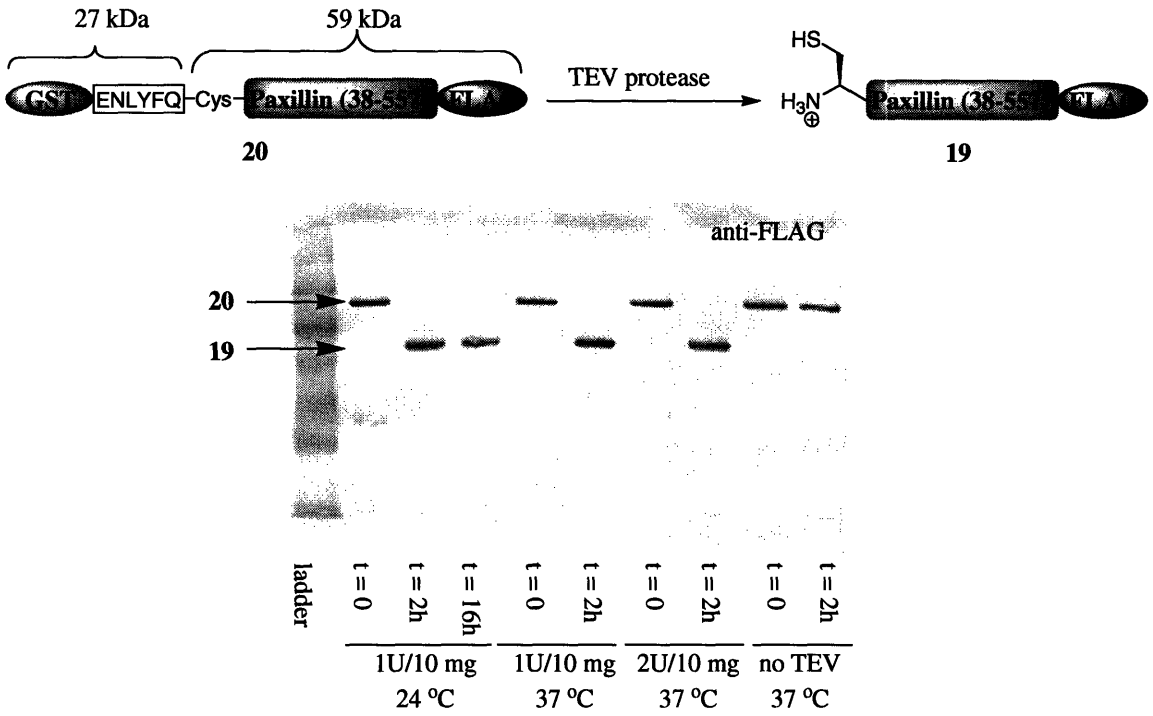


Figure 3.9. TEV protease cleavage of GST-ENLYFQ-Cys-Pax(38-557)-FLAG (**20**). A Western blot shows the protein immediately prior to protease addition (t = 0) and 2 hours after addition (t = 2 h) in various reaction conditions. A control with no TEV is also shown.

3-5. Native Chemical Ligation

In general, NCL reactions were carried out with 50 μM protein, 0.8 mM peptide, and 150 mM MESNA in TBS at pH 8.0 for 24 hours at room temperature A number of conditions were screened to determine this optimized ligation protocol. The effect of the thiol additive, pH, and peptide and protein concentrations are discussed below, as is the inhibitory effect of glycerol.

Effect of glycerol on NCL

The expressed and TEV-cleaved *C*-terminal fragment of paxillin, **19**, was stored in aliquots with 20% glycerol at -20 °C, since paxillin rapidly degrades at 4 °C, having a shelf life of less than two weeks. Buffer exchange into TBS removed most of the glycerol prior to ligation. Initial attempts to ligate **19** to the synthetic *N*-terminal peptide thioesters **18a** and **18b** failed to produce significant amounts of ligation product, despite an extensive investigation of reaction conditions. The reactions were attempted using a

variety of thiol additives, pH ranges, temperatures, and concentrations of both the protein and peptide fragments. Surprisingly, the only conditions that resulted in any detectable ligation product, although still with very poor conversion, were those reactions carried out at low pH (6.0 to 6.5). This was a particularly confusing result, since NCL is expected to work best at higher pH, where thiols are more nucleophillic.

As was later suggested in an e-mail correspondence by Professor Tom Muir, the reaction was stunted in all cases, regardless of other reaction conditions, by the exposure of the protein solution to glycerol. While glycerol itself is not problematic, commercial glycerol is contaminated with small amounts of glyceraldhyde. The glyceraldehyde essentially caps the free cysteine residues by reacting to give a thiazolidine derivative (**Fig. 3.10**). Since acid shifts the equilibrium towards the free cysteine thiol, it is logical that the only reaction conditions that resulted in any detectable ligation product were at or below pH 6.5. This capping complication was easily avoided by exposing the protein to glycerol only after NCL was complete, meaning paxillin was freshly lysed, purified, and treated with TEV immediately prior to each ligation. The incompatibility of NCL with glycerol is not mentioned in any published reports, possibly because the majority of NCL examples in the literature are on robust protein fragments that do not require freezing for short term storage.

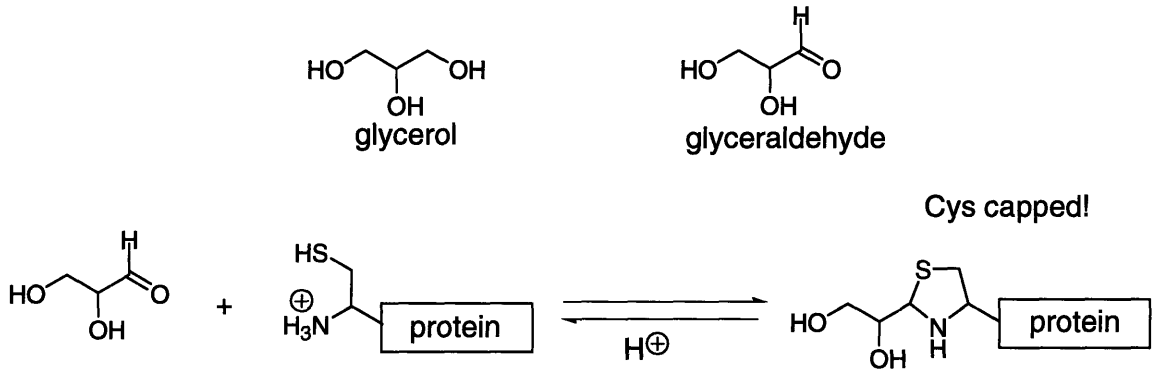


Figure 3.10. The reaction of an *N*-terminal cysteine residue with glyceraldehyde to form a thiazolidine analog. The protein cannot participate in NCL in the capped form.

Effect of thiol additive, pH and protein/peptide concentrations

Attempts to catalyze the reaction using 1 to 2% (vol/vol) thiophenol, benzylmercaptan, or a combination of the two thiols resulted in precipitate formation and

no detectable ligation product. While these thiols are ideal additives for peptide ligations, which can tolerate the addition of organic solvent, their poor water solubility makes them incompatible with ligations of full-length proteins. In contrast, MESNA (**Fig. 1.2c**) is a highly water soluble thiol and was found to successfully promote the ligations at concentrations ranging from 100 to 200 mM. MPPA, a thiol recently identified as an ideal NCL catalyst[24] (see Chapter 1 for a complete discussion of thiol additives in NCL) was also tested, but was not found to increase the ligation conversion over MESNA. In addition, MPPA was found to have low solubility in ligation conditions.

The effect of pH on ligation efficiency was briefly investigated with the reaction compared at pH 6.1 and pH 8.0 (**Table 1**). Precipitation occurred at pH 6.1 both with MESNA and 0.2% thiophenol as the thiol additive. A small amount of ligation product (~ 10% conversion) was visible in each case in the precipitate. As expected, the ligation was significantly more successful in identical conditions at pH 8.0 (~ 80% conversion). The ligation was also tested at pH 7.5 and 7.0 and found to have approximately 70% and 50% conversion, respectively.

The effects of protein, MESNA, and thioester concentration were also examined. Increasing by a factor of two either the protein concentration, from 2.5 to 5 mg/mL, or the MESNA concentration, from 100 to 200 mM, had no effect on ligation efficiency (**Table 1, g-i**). Additionally, the ligation was run with thioester concentrations of 0.65, 1, 2, 4, and 8 mM (**Table 1, j-n**) with no apparent effect on the reaction when the concentration was at, or under, 4 mM. With 8 mM thioester, the lyophilized peptide never fully dissolved and the ligation conditions resulted in a large amount of precipitation after 24 h. Since increasing the excess of thioester did not improve the ligation efficiency, future ligations were carried out at 0.8 mM thioester.

Table 1 Summary of EPL reactions with Cys-Pax(38-557)-FLAG (**19**). All reactions were carried out with 50% v/v 100 mM Tris, 600 mM NaCl, pH 8.0 or 100 mM MES, 600 mM NaCl, pH 6.1.

	thioester	rxn. size	protein amt (mg/mL)	thioester mg/mL	thiol additive	pH	precip.	product
a	pY31	30 μL	2.2	2.5 (0.5 mM)	100 mM MESNA	8.0	no	yes (60%)
b	pY31	20 μL	1.7	2.5 (0.5 mM)	100 mM MESNA	6.1	yes	in pellet (10%)
c	pY31	20 μL	1.7	2.5	0.2%	6.1	yes	in pellet

				(0.5 mM)	thiophenol			
d	Y31	20 µL	2	3.75 (0.78 mM)	100 mM MESNA	8.0	no	yes (50%)
e	Y31	20 µL	2	3.75 (0.78 mM)	100 mM MESNA	6.1	yes	in pellet (10%)
f	Y31	20 µL	2	3.75 (0.78 mM)	0.2% thiophenol	6.1	yes	in pellet
g	Y31	20 µL	2.5	3.1 (0.65 mM)	100 mM MESNA	8.0	no	yes (80%)
h	Y31	20 µL	2.5	3.1 (0.65 mM)	200 mM MESNA	8.0	no	yes (80%)
i	Y31	20 µL	5	3.1 (0.65 mM)	200mM MESNA	8.0	no	yes (80%)
j	Y31	20 µL	2.5	5.0 (1.0 mM)	100 mM MESNA	8.0	no	yes (80%)
k	Y31	20 µL	2.5	10 (2.0 mM)	100 mM MESNA	8.0	no	yes (80%)
l	Y31	20 µL	2.5	20 (4.0 mM)	100 mM MESNA	8.0	no	yes (80%)
m	Y31	20 µL	2.5	40 (8.0 mM)	100 mM MESNA	8.0	yes	in pellet (50%)
n	Y31	80 µL	2.5	3.1 (0.65 mM)	100 mM MESNA	8.0	no	yes (80%)
o	Y31	168 µL	2.2	7.1 (1.5 mM)	100 mM MESNA	8.0	no	yes (80%)

Using the optimized conditions, ligation of **19** with **18a**, **18b**, and **18c** proceeded with approximately 80% conversion to **Y31Pax** (**21a**), **pY31Pax** (**21b**) and **cpY31Pax** (**21c**) respectively. The *N*-terminal hexahistidine tag, introduced with the thioester peptides, provided a handle for purification of the final ligation product from any unligated **19** or uncleaved **20** remaining in the crude ligation mixture. Excess peptide (~5 kDa) was removed by buffer exchange using 50-kDa MWCO centrifugal filters. Gel analysis of the step-by-step semisynthesis of **cpY31Pax** (**21c**) is shown in **Figure 3.11**.

Conclusion

Here we report the semisynthesis of paxillin analogs using an NCL strategy to introduce *N*-terminal peptide thioesters, allowing us to access tens of milligrams of cpY31Pax (**21c**), pY31Pax (**21b**) and Y31Pax (**21a**). The semisynthetic strategy detailed demonstrates a straightforward approach for assembling complex eukaryotic proteins via NCL.

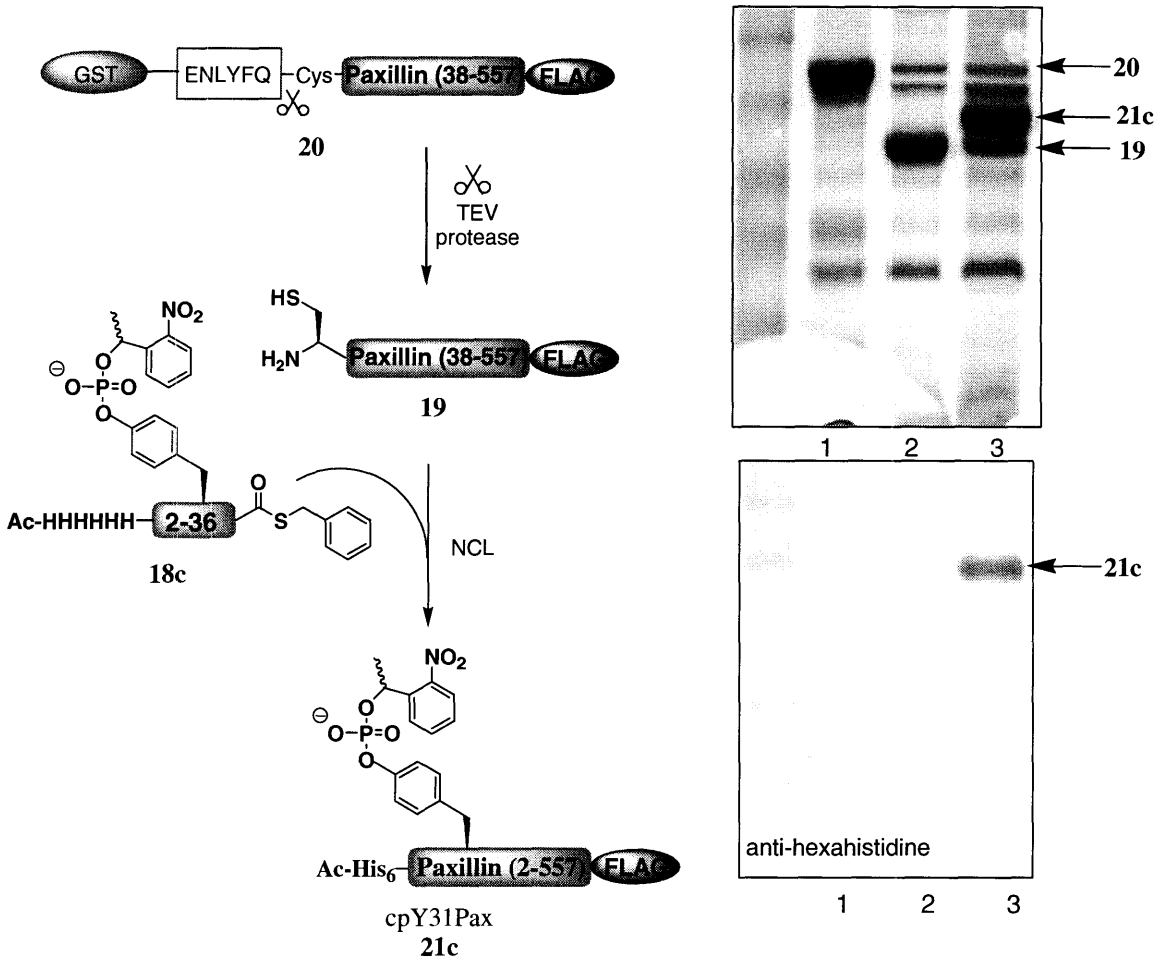


Figure 3.11. Semisynthesis of caged phosphorylated paxillin, **21c**. Coomassie and Western blot analysis of (lane 1) the purified GST-fusion protein, 20, (lane 2) the TEV cleavage product mixture, and (lane 3) the NCL product mixture.

The prevalence of reports of semisynthetic proteins modified on the *C*-terminus compared to those modified at the *N*-terminus has been attributed to the challenges of thioester generation following SPPS.[25] The method we describe is compatible with caged phosphoamino acids and, combined with the use of an Fmoc-protected caged phosphoamino acid, requires only a single reaction (thioesterification) beyond standard SPPS. Importantly, this design allows the incorporation of a variety of probes at the *N*-terminus of paxillin. Work is underway toward the construction of a paxillin variant containing an environment-sensitive fluorophore as a reporter of SH2-domain binding (with Matthieu Sainlos). Cellular uncaging followed by a fluorescent reporter for binding

will allow for time-specific information on the molecular events following phosphorylation.

The semisynthetic caged phosphoprotein **cpY31Pax** (**21c**) can be used to probe the role of Tyr31 paxillin phosphorylation in the context of the entire paxillin protein with temporal and spatial control in a cell. The in vitro validation of the binding interactions and uncaging properties of the probe are detailed in Chapter 4.

Methods

Paxillin(2-26) thioester synthesis (**19a**, **19b**, **19c**). Peptides were synthesized manually and on an automated peptide synthesizer (Advanced ChemTech 396) using standard Fmoc-based SPPS protocols on preloaded Fmoc-Gly-Novasyn TGT resin (Novabiochem). Phosphotyrosine was introduced as Fmoc-Tyr(PO(OBzl)OH)-OH, and caged phosphotyrosine was introduced (manually) as Fmoc-phospho(1-nitrophenylethyl-2-cyanoethyl)-threonine. Residues Ile18, His19, and the final four histidine residues of the hexahistidine tag were always coupled manually. For acetylation of the amino-terminus of each peptide, 120 μmol of peptide on resin were treated with acetic anhydride (113 μL, 1.2 mmol) and pyridine (97 μL, 1.2 mmol) in 10 mL of DMF for 30 min. The peptides were individually cleaved from the resin with side-chain protection intact by agitating with 0.5% TFA in DCM for 1.5 h. The resin was removed by filtration and rinsed with DCM, the solvent was mostly evaporated under a stream of nitrogen, and the peptide was triturated with cold hexanes. The hexanes were removed in vacuo, and the resulting white powder was dissolved in anhydrous THF (50 mL) and treated with HBTU (180 mg, 480 μmol), DIPEA (165 μL, 960 μmol), and benzylmercaptan (54 μL, 480 μmol) under nitrogen overnight. The THF was removed in vacuo and the peptide was deprotected with 95% (vol/vol) TFA, 2.5% (vol/vol) TIS, and 2.5% (vol/vol) water for 2 h. Peptides were triturated with cold diethyl ether, and purified by reverse phase HPLC on a Waters 600 instrument with a YMC C_{18} preparative column, using an elution gradient of water/acetonitrile with 0.1% TFA. The identities of the peptides as free acids and of the final peptide thioester products were confirmed by electrospray ionization

(ESI) mass spectrometry on a Perspective Biosystems *Mariner Biospectrometry Workstation* (turbo ion source).

Peptide characterization.

Ac-HHHHHHDDLDALLADLESTTSHISKRPVFLSEETP-Y-SYPTG-COOH, Reverse phase HPLC (t_R = 25.4 min). Exact mass calcd for $C_{209}H_{307}N_{59}O_{68}$, 4731.2; found by MS(ESI), 947.8 $[M5H]^{5+}$, 790.0 $[M6H]^{6+}$.

Ac-HHHHHHDDLDALLADLESTTSHISKRPVFLSEETP-Y-SYPTG-COSBn (**19a**), Reverse phase HPLC (t_R = 25.7 min). Exact mass calcd for $C_{216}H_{313}N_{59}O_{67}S$, 4837.3; found by MS(ESI), 968.7 $[M5H]^{5+}$.

Ac-HHHHHHDDLDALLADLESTTSHISKRPVFLSEETP-pY-SYPTG-COOH, Reverse phase HPLC (t_R = 25.4 min). Exact mass calcd for $C_{209}H_{308}N_{59}O_{71}P$, 4811.2; found by MS(ESI), 1204.6 $[M4H]^{4+}$, 963.9 $[M5H]^{5+}$.

Ac-HHHHHHDDLDALLADLESTTSHISKRPVFLSEETP-pY-SYPTG-COSBn (**19b**), Reverse phase HPLC (t_R = 25.6 min). Exact mass cald for $C_{216}H_{314}N_{59}O_{70}PS$, 4917.2; found by MS(ESI), 1231.9 $[M4H]^{4+}$, 985.7 $[M5H]^{5+}$.

Ac-HHHHHHDDLDALLADLESTTSHISKRPVFLSEETP-cpY-SYPTG-COOH, Reverse phase HPLC (t_R = 25.4 min). Exact mass calcd for $C_{217}H_{315}N_{60}O_{73}P$, 4960.3; found by MS(ESI), 993.1 $[M5H]^{5+}$.

Ac-HHHHHHDDLDALLADLESTTSHISKRPVFLSEETP-cpY-SYPTG-COSBn (**19c**), Reverse phase HPLC (t_R = 25.7 min). Exact mass calcd for $C_{224}H_{321}N_{60}O_{72}PS$, 5066.3; found by MS(ESI), 1014.9 $[M5H]^{5+}$, 845.9 $[M6H]^{6+}$.

Plasmid construction for GST-paxillin(38-557)-FLAG. The gene fragment encoding residues 38-557 of paxillin (isoform α) was amplified from a paxillin plasmid (supplied by Professor Martin Schwartz) with primers to insert 5'-EcoR1 and 3'-Not1 restriction sites. The primers also encoded an amino-terminal TEV protease recognition site (ENLYGQC) and a carboxy-terminal FLAG tag. For this amplification the following PCR primers were used: 5'- GCC GGA ATT CGT GAA AAC CTG TAT TTT CAG TGC CAC ACA TAC CAG GAG ATT -3' and 5'- GCC CCC TTT TGC GGC CGC CTA CTT ATC GTC ATC GTC TTT GTA GTC GCA GAA GAG CTT GAG GAA-3'. The PCR-amplified

fragments were digested with Not1 and EcoR1 and ligated to Not1/EcoR1-digested and CIP-treated pGEX-4T-2 (GE Health Sciences). The ligation mixture was transformed into DH5α cells and grown on carbenicillin-resistant plates. Plasmid DNA was isolated from selected colonies and confirmed by sequencing.

GST-paxillin(38-557)-FLAG expression and purification. The paxillin plasmid was transformed into BL21-CodonPlus®-RP competent cells (Stratagene) and grown with fermentation at 37 °C to midlog phase in 10 L of LB media with carbenicillin and chloramphenicol. The culture was cooled to 16 °C, and the cells were induced with 0.1 mM IPTG and fermented for 20 to 24 h. Cells were harvested by centrifugation (4.2 k rpm for 30 min), and the resulting pellet was frozen at -80 °C. For cell lysis, pellets were thawed and resuspended in 400 mL of lysis buffer (PBS, 1 mM $ZnCl_2$, 1 mg/mL lysozyme, 1 mM DTT, and 0.1% (vol/vol) of protease inhibitor cocktail III [100 μM AEBSF, 80 nM aprotinin, 5 μM bestatin, 1.5 μM E-64, 2 μM leupeptin, 1 μM pepstatin A] (Calbiochem)) and incubated for 20 min at 4 °C. NP-40 Alternative (1% vol/vol, 4 mL per 400 mL of lysis buffer) was added to the resuspended cells, and the mixture was agitated at rt for 10m. The lysed cells were sonicated on ice for 3 min at 50% power and 30% duty and then subjected to centrifugation (13,000 rpm for 30 min, then 35,000 rpm for 30 min).

The soluble fraction was purified using Glutathione Sepharose 4 Fast Flow resin (GE Health Sciences) following the manufacturer's protocol. Accordingly, the protein solution was stirred with GST-affinity resin (8 mL) for approximately 1 h at 4 °C. The resin was washed with PBS (3 x 100 mL), and the GST-tagged protein was eluted by incubation with elution buffer (50 mM Tris, pH 8, 10 mM reduced glutathione, and 0.1% (vol/vol) protease inhibitor cocktail III (Calbiochem)) for 3 min at °4 C, followed by collection of the flow through. For a 10-L preparation, three 8-mL elution fractions were collected and combined. The protein was dialyzed into TBS (3 x 2 L) using 50-mm flat width spectra/por membrane (Spectrum), and then purified via the carboxy-terminal tag with anti-FLAG M2 affinity resin (Sigma) as follows. The dialyzed protein was incubated with FLAG-affinity resin (3 mL) with gentle agitation at 4 °C, and the resin was washed with TBS (3 x 30 mL). FLAG-labeled protein was eluted with 1.2-mL

aliquots of 100 mM glycine·HCl, pH 3.5. To each elution was immediately added 160 μL of 500 mM Tris pH 8.0, and 100 uL of 10 x TBS to neutralize and add salt to the solution. Typical yields for the doubly-purified protein were 4 to 6 mg per 10-L fermentation, as quantified using a Biorad protein assay. The protein was analyzed by 10% SDS-PAGE gels, and visualized with Coomassie blue dye and by Western blot with a mouse anti-FLAG primary antibody.

TEV protease cleavage. The purified protein GST-ENLYFQC-Paxilllin(38-557)-FLAG (**21**) was diluted to 0.5 or 1 mg/mL into a TEV cleavage buffer with a final concentration of 50 mM Tris pH 8.0, 500 μM EDTA, and 5 mM BME. TEV protease (US Biological) was added (35 μL of protease per mg of protein), and the resulting solution was incubated at 27 °C for 3 h. The protein was dialyzed into TBS (to remove any glycine present from the FLAG-affinity elution) and incubated with Ni/NTA resin and glutathione sepharose beads to remove the hexahistidine-tagged TEV protease and the cleaved GST tag. The flow-through was concentrated using 50 KDa MWCO centrifugal filters (Millipore) and used immediately in NCL.

Ligations. In general, reactions were carried out with 50 μM protein, 0.8 mM peptide, and 100 mM MESNA in TBS at pH 8.0. Accordingly, to a solution of Cys-Pax(38-557)-FLAG (**21**) (600 μg, 10.7 nmol) in TBS (150 μL) was added Ac-HHHHHH-DDLDALLADLESTTSHISKRPVFLSEETP-X-SYPTG-COSBn (lyophilized, then dissolved into 20 μL of water for transfer) (800 μg, 169 nmol for X = Tyr (**19a**), 163 nmol for X = pTyr (**19b**), and 158 nmol for X = cpTyr (**19c**)), 10 μL of 2 M MESNA, and 20 μL of 500 mM Tris pH 8.0. The reaction was incubated for 20 h at 25 °C, then dialyzed into PBS using a 50-kDa MWCO dialysis membrane to remove excess (4.8 to 5.1 kDa) peptide. Protein was either used directly for assays without a final purification, or a purified using a Ni/NTA spin column to isolate ligation product via the ligation-introduced amino-terminal hexahistidine tag. The protein was analyzed by 10% SDS-PAGE gels, and visualized with Coomassie blue dye, and by Western blot with a mouse anti-hexahistidine primary antibody. For ligations using **19b**, a mouse anti-pY31 antibody was also used for visualization. The purified protein was stored at 4 °C and

used for all in vitro and cellular studies within 2 weeks of lysis and purification. For longer term storage (up to several months), the protein was stored at 1 mg/mL in PBS with 20% glycerol at -20 °C.

Acknowledgements: We thank Professor Martin Schwartz for donating the paxillin construct.

References

1. Rothman, D. M., Vazquez, M. E., Vogel, E. M. & Imperiali, B. Caged Phospho-Amino Acid Building Blocks for Solid-Phase Peptide Synthesis. *J. Org. Chem.* **68**, 6795-6798 (2003).
2. Rothman, D. M., Shults, M. D. & Imperiali, B. Chemical approaches for investigating phosphorylation in signal transduction networks. *Trends Cell Biol.* **15**, 502-510 (2005).
3. Turner, C. E. Paxillin and focal adhesion signalling. *Nat. Cell Biol.* **2**, E231-E236 (2000).
4. Lauffenburger, D. A. & Horwitz, A. F. Cell migration: a physically integrated molecular process. *Cell (Cambridge, Mass.)* **84**, 359-69 (1996).
5. Matthews, C. K. & van Holde, K. E. *Biochemistry* (The Benjamin/Cummings Publishing Company, Menlo Park, 1995).
6. Brown, M. C. & Turner, C. E. Paxillin: Adapting to change. *Physiol. Rev.* **84**, 1315-1339 (2004).
7. Schaller, M. D. & Parsons, J. T. pp125FAK-dependent tyrosine phosphorylation of paxillin creates a high-affinity binding site for Crk. *Mol. Cell. Biol.* **15**, 2635-45 (1995).
8. Brown, M. C., Curtis, M. S. & Turner, C. E. Paxillin LD motifs may define a new family of protein recognition domains. *Nat. Struct. Biol.* **5**, 677-678 (1998).
9. Brown, M. C. & Turner, C. E. Roles for the tubulin- and PTP-PEST-binding paxillin LIM domains in cell adhesion and motility. *Int. J. Biochem. Cell Biol.* **34**, 855-863 (2002).

10. Shen, Y., Schneider, G., Cloutier, J.-F., Veillette, A. & Schaller, M. D. Direct association of protein-tyrosine phosphatase PTP-PEST with paxillin. *J. Biol. Chem.* **273**, 6474-6481 (1998).

11. Brown, M. C., Perrotta, J. A. & Turner, C. E. Serine and threonine phosphorylation of the paxillin LIM domains regulates paxillin focal adhesion localization and cell adhesion to fibronectin. *Mol. Biol. Cell* **9**, 1803-1816 (1998).

12. Petit, V. et al. Phosphorylation of tyrosine residues 31 and 118 on paxillin regulates cell migration through an association with CRK in NBT-II cells. *J. Cell Biol.* **148**, 957-969 (2000).

13. Vindis, C., Teli, T., Cerretti, D. P., Turner, C. E. & Huynh-Do, U. EphB1-mediated Cell Migration Requires the Phosphorylation of Paxillin at Tyr-31/Tyr-118. *J. Biol. Chem.* **279**, 27965-27970 (2004).

14. Jamieson, J. S. et al. Paxillin is essential for PTP-PEST-dependent regulation of cell spreading and motility: A role for paxillin kinase linker. *J. Cell Sci.* **118**, 5835-5847 (2005).

15. Vande Pol, S. B., Brown, M. C. & Turner, C. E. Association of bovine papillomavirus type 1 E6 oncoprotein with the focal adhesion protein paxillin through a conserved protein interaction motif. *Oncogene* **16**, 43-52 (1998).

16. Turner, C. E. Paxillin. *Int. J. Biochem. Cell Biol.* **30**, 955-959 (1998).

17. Iwasaki, T. et al. Involvement of phosphorylation of Tyr-31 and Tyr-118 of paxillin in MM1 cancer cell migration. *Int. J. Cancer* **97**, 330-335 (2002).

18. Hackeng, T. M., Griffin, J. H. & Dawson, P. E. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10068-10073 (1999).

19. Mezo, A. R., Cheng, R. P. & Imperiali, B. Oligomerization of uniquely folded mini-protein motifs: development of a homotrimeric $\beta\beta\alpha$ peptide. *J. Am. Chem. Soc.* **123**, 3885-3891 (2001).

20. Futaki, S., Sogawa, K., Maruyama, J., Asahara, T. & Niwa, M. Preparation of peptide thioesters using Fmoc-solid-phase peptide synthesis and its application to the construction of a template-assembled synthetic protein (TASP). *Tetrahedron Lett.* **38**, 6237-6240 (1997).

21. Erlanson, D. A., Chytil, M. & Verdine, G. L. The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex. *Chem. Biol.* **3**, 981-991 (1996).

22. Tolbert, T. J. & Wong, C.-H. New methods for proteomic research: preparation of proteins with N-terminal cysteines for labeling and conjugation. *Angew. Chem., Int. Ed. Engl.* **41**, 2171-2174 (2002).

23. Eaton, D., Rodriguez, H. & Vehar, G. A. Proteolytic processing of human factor VIII. Correlation of specific cleavages by thrombin, factor Xa, and activated protein C with activation and inactivation of factor VIII coagulant activity. *Biochemistry* **25**, 505-12 (1986).

24. Johnson, E. C. B. & Kent, S. B. H. Insights into the mechanism and catalysis of the native chemical ligation reaction. *J. Am. Chem. Soc.* **128**, 6640-6646 (2006).

25. Schwarzer, D. & Cole Philip, A. Protein semisynthesis and expressed protein ligation: chasing a protein's tail. *Curr. Opin. Chem. Biol.* **9**, 561-9 (2005).

Chapter 4

In vitro characterization of semisynthetic paxillin analogs and preliminary cellular studies on adhesion dynamics

Introduction

The properties of the semisynthetic proteins described in Chapter 3, Y31Pax (**21a**), pY31Pax (**21b**), and cpY31Pax (**21c**) (**Fig. 4.1**), were characterized through in vitro binding, phosphorylation, and uncaging studies. Following the in vitro analyses, the caged phosphoprotein probe **21c** was applied in preliminary cellular studies to examine the influence of Tyr31 paxillin phosphorylation on focal adhesion dynamics in cellular migration.

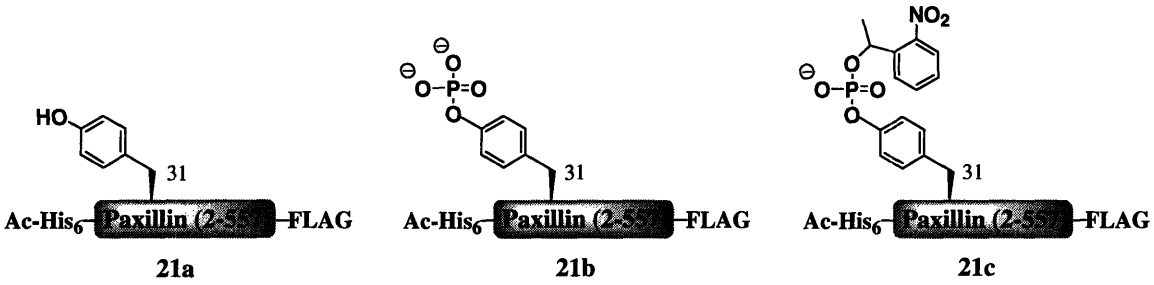


Figure 4.1. Semisynthetic paxillin analogs.

In vitro characterization of paxillin analogs

Since paxillin is a molecular adaptor protein with no known enzymatic activity, direct validation of the functionality of the semisynthetic paxillin derivatives is not possible by observing reaction turnover. Instead, we examined the properties of the reconstituted paxillin control, Y31Pax (**21a**), in vitro by analyzing the ability to bind known paxillin binding partners and act as a substrate for identified paxillin kinases. Paxillin-interacting proteins focal adhesion kinase (FAK), GRK interactor 1 (GIT1), and PTP-PEST were selected to probe binding along the entire length of paxillin. In addition, the substrate recognition of Y31Pax (**21a**) by native paxillin kinases was probed with phosphorylation assays focused on the NCL-introduced *N*-terminus of paxillin, including the Tyr31 phosphorylation site (**Fig. 4.2**). These assays examined the phosphorylation of paxillin by Src, ERK, and JNK kinases. The binding and phosphorylation experiments

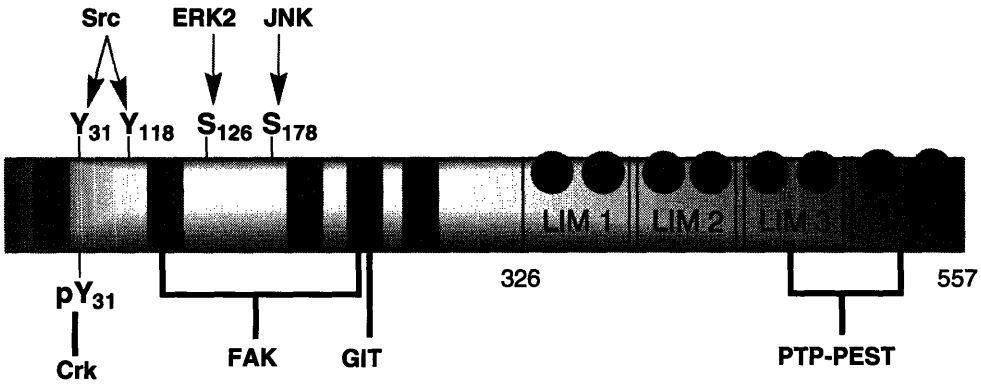


Figure 4.2. Paxillin with selected binding partners and upstream kinases. Binding interactions are depicted with solid black lines and phosphorylation is indicated with red arrows.

confirmed that the semisynthetic reconstituted control, Y31Pax (**21a**), behaved as expected, and the function of paxillin was not compromised by the expression or ligation procedures.

The extent of cpY31Pax (**21c**) uncaging was quantified following irradiation with light centered at 365 nm. Close to 90% conversion of **21c** to pY31Pax (**21b**) resulted from 90 seconds of irradiation on a standard DNA transilluminator. The caged phosphoprotein probe was then subjected to Crk binding prior to and following uncaging. The Crk SH2 domain is known to bind paxillin at a site that includes phosphoTyr31, and the results of the binding assay suggest that **21c** effectively exposed a CrkSH2 binding site after irradiation, resulting in up to a three-fold increase in CrkSH2 binding.

Preliminary cellular focal adhesion experiments

Following the in vitro experiments, cpY31Pax (**21c**) was used to examine the role of Tyr31 paxillin phosphorylation on focal adhesion turnover in paxillin-null murine embryonic fibroblast (MEF) cells. Preliminary experiments were run over a two week period in the lab of Professor Rick Horwitz (University of Virginia) to assess the potential of a full investigation and determine the necessary experimental plan and controls.

Focal adhesions are sites of cellular contact with the extracellular matrix that facilitate cellular adhesion, migration, and signaling. Adhesions form a physical link between the external environment and the cytoplasm of a cell, which enables a cell to

respond to extracellular signals, such as growth factors. Structurally, these cell-matrix contacts are composed of both structural and signaling proteins, and they include clusters of integrins, which are heterodimeric transmembrane proteins.[1] Focal adhesions provide both rigidity, to enable strong adhesion to the matrix, and plasticity, with the capacity for rapid formation and disassembly (turnover), to facilitate dynamic cellular responses to extracellular signals.[2] In migrating cells, continuous turnover of focal adhesions occurs at the leading edge of a cell, and concomitant adhesion disassembly occurs at a cell's following edge (**Fig. 4.3**). Adhesion dynamics are tightly regulated, but the process is currently not well understood and can be difficult to study.[3] As discussed in Chapter 3, paxillin is an adaptor protein that localizes to focal adhesions and mediates protein-protein signaling interactions. The phosphorylation of paxillin at specific residues creates protein binding sites and is implicated in the controlled assembly and dissolution of signaling cascades.[4] It has been found that paxillin-null fibroblasts contain abnormal focal adhesions, suggesting a role for paxillin in adhesion maintenance or turnover.[5]

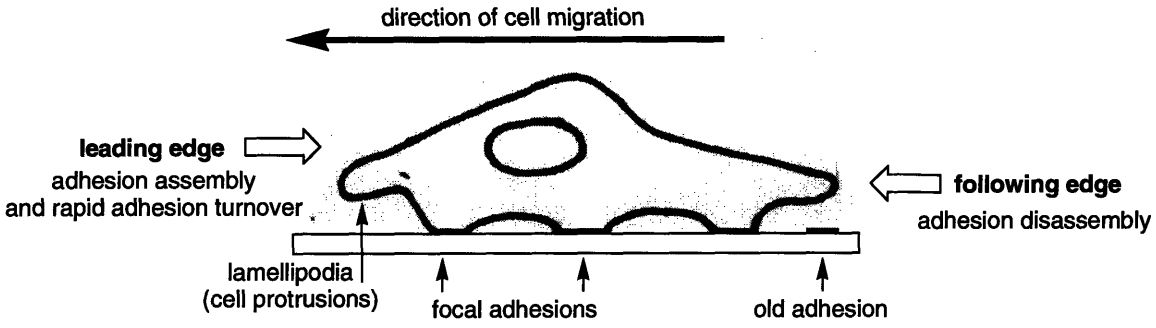


Figure 4.3. Cartoon of a migrating cell. Adhesions (contacts with the extracellular matrix) are indicated in red.

In this chapter, we describe the cloning and expression of paxillin-binding fragments of FAK, GIT1, and PTP-PEST and the results of subsequent binding, phosphorylation, uncaging, and Crk-association experiments. We also report the promising results of preliminary focal adhesion turnover experiments using **21c** to probe the effect of Tyr31 paxillin phosphorylation on adhesion dynamics.

Results and Discussion

4-1. Paxillin binding assays

FAK and GIT are proteins known to bind to the *N*-terminal LD repeats of paxillin. FAK, a protein tyrosine kinase, constitutively binds paxillin in vivo, interacting with the LD2 and LD4 motifs.[4] The paxillin-binding region of FAK has been identified by in vitro deletion analysis to be in the FAK *C*-terminal domain, within residues 904-1054.[6] GIT interacts with the LD4 motif of paxillin, and the 140 *C*-terminal amino acids of GIT have been shown to be sufficient for paxillin binding.[7] PTP-PEST, a phosphatase, binds directly to the *C*-terminal LIM3 and LIM4 domains of paxillin (**Fig. 4.2**).[4] The non-catalytic *C*-terminal domain of PTP-PEST has been shown to interact with paxillin in vitro and in vivo, although the *N*-terminal 300 residues containing the catalytic domain have no paxillin association. Deletion constructs have been used to identify the location of paxillin binding on PTP-PEST to be within residues 300-494.[8]

For the binding studies presented herein, *N*-terminal GST tags, which are absent in the ligated paxillin constructs, were incorporated onto paxillin-binding regions of FAK, GIT, and PTP-PEST. The three GST-tagged constructs, GST-FAK(857-1057), GST-GIT(622-761), and GST-PTP-PEST(338-390), were tested for binding with the semisynthetic Y31Pax (**21a**) in a modified GST pull-down assay.

Preparation of GST-FAK(857-1053), GST-GIT(622-761), and GST-PTP-PEST(338-390)

Inserts containing the paxillin-interacting regions of FAK and GIT were prepared for cloning into pGEX-4T-2 GST-fusion vectors using the appropriate primers. After PCR and cloning, the correct sequences of the isolated plasmid DNA were confirmed by sequencing, and the vectors were transformed into BL21-CodonPlus®-RP competent cells for expression of GST-FAK(857-1053) and GST-GIT(622-761). The RP-enhanced cells, which encode the AGG (Arg), AGA (Arg), and CCC (Pro) tRNAs, were selected to compensate for the Arg and Pro rare codons present in the FAK and GIT constructs. The paxillin-interacting region of PTP-PEST was supplied by Professor Mike Schaller in a vector containing an *N*-terminal GST fusion to residues 338-390. Isolated DNA from this vector and an empty pGEX-4T-2 vector were transformed into BL21 CodonPlus®-RIL cells to produce GST-PTP-PEST(338-390) and GST respectively. Protein expression for

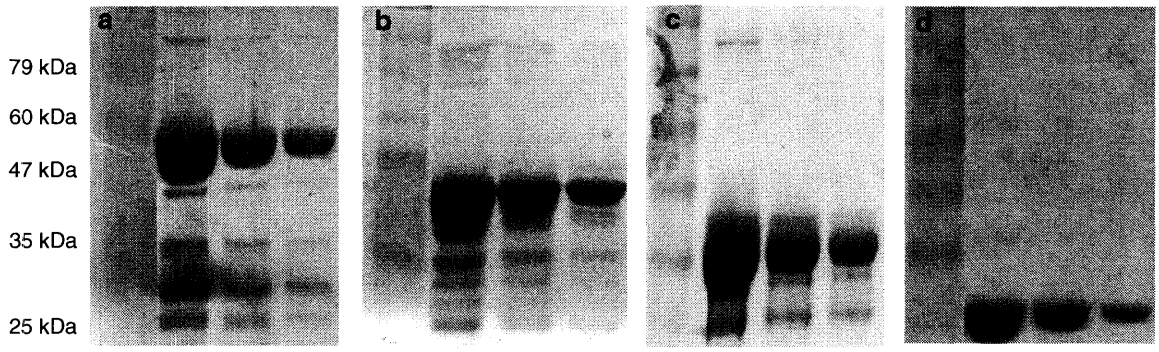


Figure 4.4. Elutions from GST-affinity purification of **a**) GST-FAK(857-1053) (49 kDa), **b**) GST-GIT(622-761) (43 kDa), **c**) GST-PTP-PEST(338-390) (33 kDa), and **d**) GST alone (27 kDa).

all constructs was carried out in LB with the appropriate antibiotics and induced at an OD_{600} of 0.6-0.8 with 0.1 mM IPTG followed by four hours of shaking at 37 °C.

The proteins were isolated from their corresponding cell lysates with glutathione sepharose beads using batch purification and eluted from the resin with 50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0. Expression yields ranged from 18 to 32 mg/L for the GST-purified proteins. Coomassie gel analysis indicated the presence of GST, GST-FAK(857-1053), GST-GIT(622-761), and GST-PTP-PEST(338-390) with calculated molecular weights of 27, 49, 43, and 33 kDa, respectively (**Fig. 4.4**). The proteins were dialyzed into PBS to remove the glutathione prior to the paxillin-binding studies, which require immobilization on glutathione sepharose beads. The purified proteins were stable for several weeks when stored at 4 °C, with the exception of GST-FAK(857-1053). The

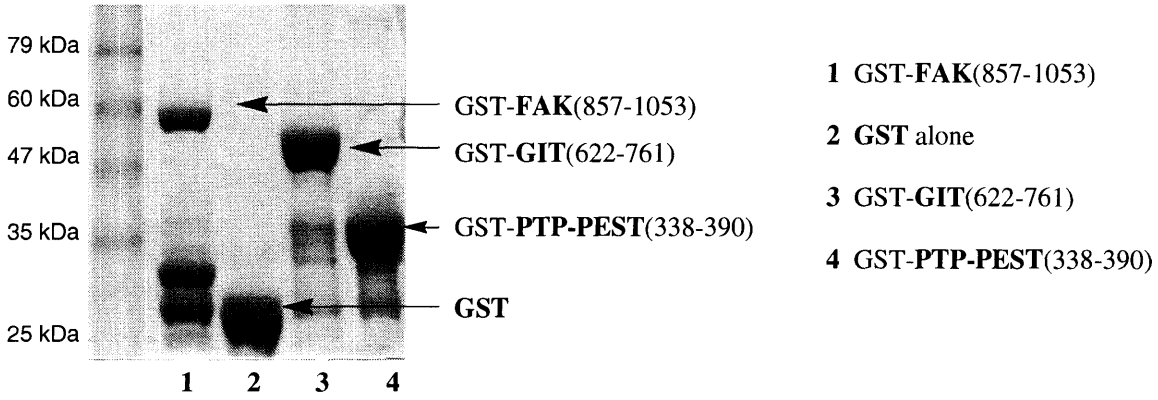


Figure 4.5. GST and GST fusion proteins, GST-FAK(857-1053), GST-GIT(622-761), and GST-PTP-PEST(338-390) in PBS, subsequent to 1 week of storage at 4 °C.

FAK construct partially degraded into two fragments over several days with the unexplained loss of the GST tag (**Fig. 4.5**). Complete degradation occurred within weeks. While there is no obvious mechanism for this cleavage, the unexpected loss of a GST tag from purified GST-fusion proteins has been observed with other GST-tagged constructs in the lab.

*Y31Pax (**21a**) binding studies*

The three GST-tagged constructs, GST-FAK(857-1057), GST-GIT(622-761), and GST-PTP-PEST(338-390), were tested for binding with the semisynthetic Y31Pax (**21a**) in a modified GST pull-down assay (**Fig. 4.6a**). The expressed GST tag alone (27 kDa) was tested as a negative control to rule out non-specific interactions. For this assay, each of the four proteins was immobilized on glutathione sepharose beads and then incubated

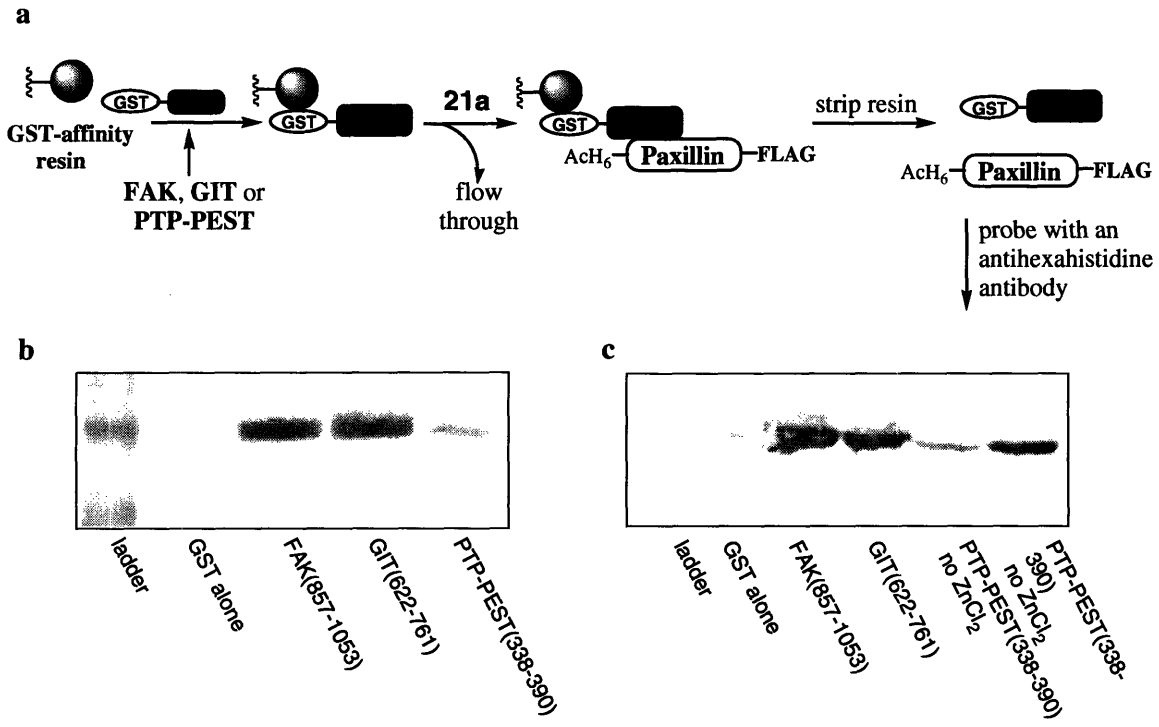


Figure 4.6. **a**) In vitro binding assay of semisynthetic paxillin (**21a**) with a negative control (GST), and GST-fusions of the paxillin binding regions of FAK, GIT, and PTP-PEST. The GST and GST fusions were bound to solid support, incubated with **21a**, washed, and then probed for paxillin binding through Western blot analysis. **b**) An anti-hexahistidine-probed Western blot of the assay in the absence of $ZnCl_2$ **c**) An anti-hexahistidine-probed Western blot of the assay in the presence of $ZnCl_2$, including one lane (as marked) of PTP-PEST incubation in the absence of $ZnCl_2$.

with **21a** in the presence of 1% BSA. The beads were copiously washed before being subjected to Western blot analysis to identify the paxillin containing samples. The Western blot was probed with an anti-hexahistidine primary antibody to detect the *N*-terminal hexahistidine tag of **21a.**

Paxillin binding was detected strongly with the FAK and GIT constructs, negligibly with GST alone, and weakly with the PTP-PEST construct (**Fig. 4.6b**). Since PTP-PEST binds at the *C*-terminal zinc-binding LIM domains of paxillin, we evaluated whether improper folding of these domains due to insufficient Zn(II) in the purified construct contributed to the poor PTP-PEST binding characteristics. The addition of $ZnCl_2$ to the paxillin solution significantly increased PTP-PEST binding (**Fig. 4.6c**). All subsequent samples of paxillin and semisynthetic paxillin derivatives were purified in the presence of 1 mM $ZnCl_2$. Through these binding experiments, we confirmed that the semisynthetic paxillin construct interacts with known paxillin binding proteins comparably to native paxillin.

4-2. In vitro paxillin phosphorylation

Next, to demonstrate that the semisynthetic paxillin analog functioned as a substrate for selected kinases that natively phosphorylate the protein, phosphorylation of **21a** was attempted with three known paxillin kinases and probed with phosphorylation specific antibodies to four sites within paxillin. Paxillin phosphorylation occurs on tyrosine, threonine, and serine residues throughout its multiple domains, and kinases that phosphorylate paxillin include Src, ERK, and JNK.[9] Src is a protein tyrosine kinase that phosphorylates paxillin on Tyr31 and Tyr118, while ERK and JNK are serine/threonine kinases that phosphorylate residues Ser126 and Ser178, respectively (**Fig. 4.7a**). As a control, and for developing assay conditions for **21a**, a full-length paxillin construct (GST-paxillin(1-557)-FLAG) was expressed and subjected to the kinase assays prior to the semisynthetic construct.

For the assays, expressed or semisynthetic paxillin was treated with Src, ERK, and JNK, and the resulting reaction mixtures were visualized by Western blots analysis using a general anti-paxillin antibody and phosphorylation-specific anti-paxillin [pY31], [pY118], [pS126], and [pS178] antibodies.

Phosphorylation assays with expressed full-length paxillin

An expressed paxillin construct was used in the kinase assays to establish appropriate in vitro phosphorylation and chemiluminescent detection conditions. Full-length paxillin (GST-Pax(1-557)-FLAG) was expressed in BL21-CodonPlus®-RP competent cells and purified by GST-affinity and FLAG-affinity columns (yield: 5 mg/10 L expression). Following a generic phosphorylation procedure, paxillin was treated with recombinant human active c-Src, ERK2, or JNK2α2 kinase (Biosource) in the presence of 15 mM $MgCl_2$ and 50 to 500 μM ATP in TBS. A negative control was run in which paxillin was subjected to the reaction conditions in the absence of a kinase. For chemiluminescent detection of paxillin, it was determined that 5 to 20 ng per well of protein was sufficient for detection by anti-paxillin and phosphospecific anti-paxillin [pY31] or [pY118] primary antibodies, using HRP-conjugated secondary antibodies. The less-sensitive anti-paxillin [pS126] and [pS178] antibodies required 200-300 ng of paxillin per well for visualization. Freshly lysed and purified paxillin was used for all phosphorylation assays.

As expected, detection of Tyr31 and Tyr118 phosphorylation in full-length paxillin was possible only after treatment with Src. Similarly, Ser126 was phosphorylated exclusively by ERK (**Fig. 4.7b**). However, Ser178 phosphorylation could not be detected after incubation with any of the kinases, including JNK2α2. DNA sequencing confirmed that the primary paxillin sequence was correct, ruling out the possibility of a mutation at Ser178 or a surrounding residue as the culprit for preventing JNK2α2 phosphorylation.

A number of conditions were explored in further attempts to achieve Ser178 phosphorylation. While JNK2α2 has been shown by Biosource to phosphorylate paxillin in vivo, no in vitro assays have yet been reported. However, JNK1 has been used in vitro to successfully phosphorylate paxillin, so those conditions, which include the addition of 1 mM DTT and 5 mM EGTA, were tested with JNK2α2. Neither the modified conditions, nor an increase in the amount of kinase by four fold resulted in detectible levels of phosphorylation. Since JNK1 phosphorylation had been described in the context of in vitro experiments, JNK1α1 (Upstate) was obtained and tested, and it was

found to phosphorylate paxillin exclusively at Ser178 (**Fig. 4.7b**). This result could be an indication of differing substrate recognition between the two isoforms of JNK or could have simply been caused by a poor commercial preparation of JNK2α2.

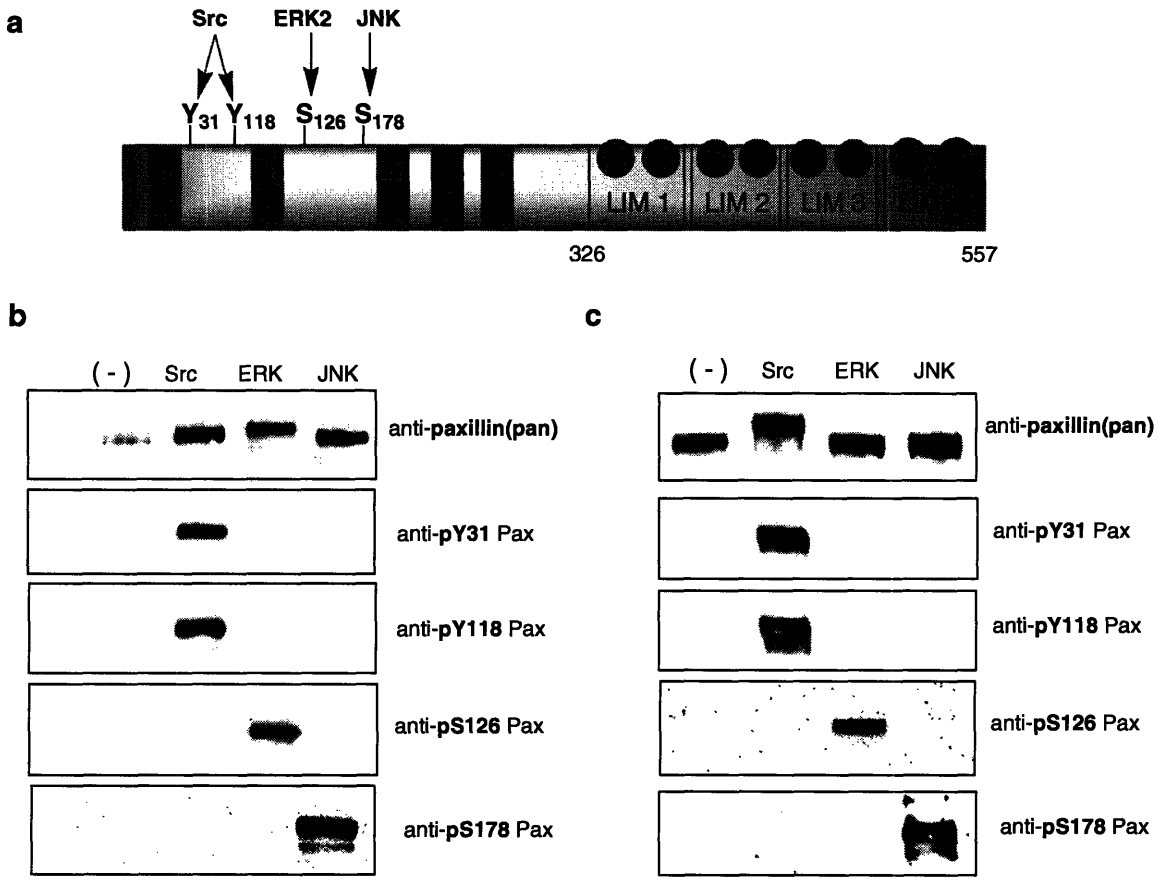


Figure 4.7. **a**) Paxillin kinases Src, ERK, and JNK. Phosphorylation sites are indicated by red arrows. Phosphorylation of **b**) expressed full-length paxillin and **c**) semisynthetic Y31Pax (**21a**) by Src, ERK, and JNK, and visualized by phospho-specific antibodies to paxillin[pY31], [pY118], [pS126], and [pS128]. The reactions were also probed with a general anti-paxillin (pan) antibody to demonstrate comparable loading. (-) indicates the negative control treated with no kinase.

Phosphorylation assays with semisynthetic paxillin

With phosphorylation and detection conditions established for Src, ERK, and JNK phosphorylation of expressed full-length paxillin, the assays were repeated using the semisynthetic paxillin construct Y31Pax (**21a**). Considering the extremely high sensitivity of chemiluminescent detection, it was encouraging that visualization of ligated

paxillin (Y31Pax) with the anti-paxillin pan antibody resulted in a clean gel containing one major band for the ligated paxillin, no detectible band for uncleaved protein (GST-ENLYFQ-Cys-Pax(38-557)-FLAG), and only a light band for unligated material (Cys-Pax(38-557) (**Fig. 4.8**).

Probing each kinase reaction with all five antibodies revealed that residues Tyr31 and Tyr118 of **21a** were phosphorylated exclusively by Src, that residue Ser126 was phosphorylated only by ERK, and that residue Ser178 was phosphorylated only by JNK (**Fig. 4.7c**). The pan paxillin antibody recognized protein in all reactions. Detection of Src-treated **21a** by the pTyr31 phosphorylation-specific antibody validated both the successful ligation to install the Tyr31 site and the recognition of that site by an upstream kinase. The identical results for the kinase assays with expressed and semisynthetic paxillin, as well as the agreement with in vivo and in vitro reports of paxillin phosphorylation, further confirmed that the semisynthetic proteins behave comparably to native paxillin.

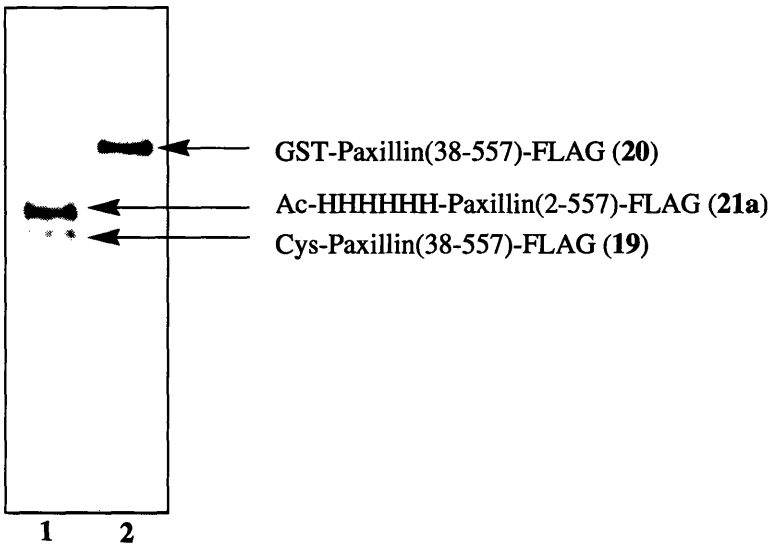


Figure 4.8. Chemiluminescent detection of lane 1) ligated paxillin, **21a**, (61 kDa) and lane 2) GST-Pax(38-557)-FLAG (**20**) (83 kDa). The blot was probed with a general mouse anti-paxillin primary antibody.

4-3. Uncaging of cpY31 paxillin

Irradiation of NPE-caged phosphopeptides with long wavelength UV light (> 365 nm) to reveal the corresponding phosphopeptides has been shown to proceed with high

quantum yields for uncaging, ranging from 0.16 to 0.5, depending on the peptide sequence.[10] Importantly for cellular applications, the photo-byproduct of NPE-caged species, *o*-nitrosoacetophenone, has previously been shown to be cellularly-inert.[11,12] This by-product is less reactive than the analogous *o*-nitrosobenzaldehyde produced by irradiation of widely-used *ortho*-nitrobenzyl-derived caging groups.[13] To ensure that cpY31Pax (**21c**) uncaging proceeded efficiently, the extent of uncaging of the ligated protein was determined following irradiation with a standard DNA transilluminator (**Fig. 4.9**). In brief, cpY31Pax (**21c**) was irradiated for 90 seconds with light centered at λ = 365 nm. The amount of phosphorylated paxillin (pY31Pax) liberated at t = 0 and t = 90 seconds post photolysis was measured by chemiluminescent detection of the phosphorylated protein. The phosphorylated species was detected using a paxillin

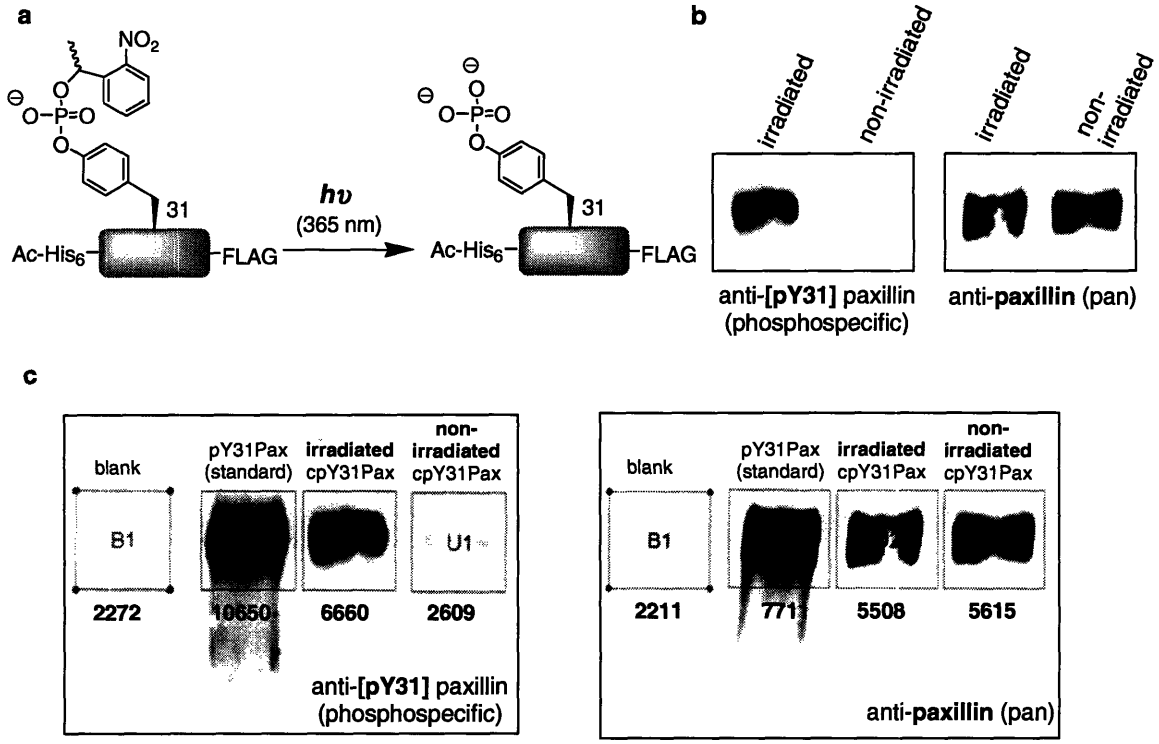


Figure 4.9. **a**) Irradiation of cpY31Pax (**21c**) with 365 nm light releases the caging group to reveal pY31Pax. **b**) Western blot analysis with a general anti-paxillin(pan) antibody and a phosphospecific anti-paxillin[pY31] antibody of **21c** prior to irradiation and following irradiation for 90 s with 365 nm light. c) To quantify uncaging, Western blots probed with anti-paxillin(pan) and anti[pY31]paxillin antibodies were visualized by chemiluminescent detection. Pixel counts are shown in bold below each lane. The gels include (B1) a blank lane, (U3) semisynthetic pY31Pax (**21b**) as a standard for pixel counts, (U2) irradiated **21c**, and (U1) non-irradiated **21c**.

[pY31] phosphorylation-specific antibody. The total amount of protein loaded on the Western blots was detected using a general anti-paxillin antibody. In addition, semisynthetic pY31 (**21b**) was used as a photochemically inert internal standard in the phosphorylation-specific and general paxillin Western blots. The relative intensity of the sample at t = 0 and t = 90 seconds was determined for each Western blot as a fraction of total pixels compared to the internal pY31Pax (**21b**) standard (**Fig. 4.9c**). The ratio of intesity between the phosphospecific and general paxillin blots was determined to calculate the extent of uncaging. The ratio of uncaged phosphopaxillin to the total protein amount indicated between 83 and 91% conversion to phosphoTyr31-paxillin. A minimal amount of uncaged phosphoprotein, typically less than 6% of the total protein, was detected in the absence of intentional irradiation using minimal precautions against UV light exposure. A sample uncaging calculation using the Western blot shown in **Fig. 4.9** is included in the Methods section.

4-4. Phosphorylation-dependant Crk binding

The SH2 domain of Crk, which recognizes a (Pro)-pTyr-X-X-Pro binding motif, has been shown to interact with paxillin following phosphorylation at Tyr31 or Tyr118 in vivo.[14,15] The effect of a concentration burst of pTyr31-paxillin on CrkSH2 binding, was investigated using the caged paxillin probe. Accordingly, the CrkSH2-binding ability of cpY31Pax (**21c**) was examined prior to and following uncaging.

As with the FAK, GIT and PTP-PEST binding studies, a CrkSH2 domain with an *N*-terminal GST tag was immobilized on glutathione sepharose beads. The beads were incubated with cpY31Pax (**21c**) with 1% BSA, either prior to or following 90 seconds of irradiation with light centered at 365 nm. The beads were washed and boiled in reducing gel loading buffer, and Western blots probed with an anti-hexahistidine primary antibody were used to identify paxillin-containing samples. Crk binding was detected up to three-fold more strongly with uncaged paxillin than with the caged construct (**Fig. 4.10**), demonstrating the ability of cpY31Pax to release the corresponding phosphorylated protein and enhance a downstream molecular interaction. Interestingly, the binding difference was less pronounced in slightly modified assay conditions. These results suggest the paxillin-Crk binding interaction is more complicated than an "on/off" event.

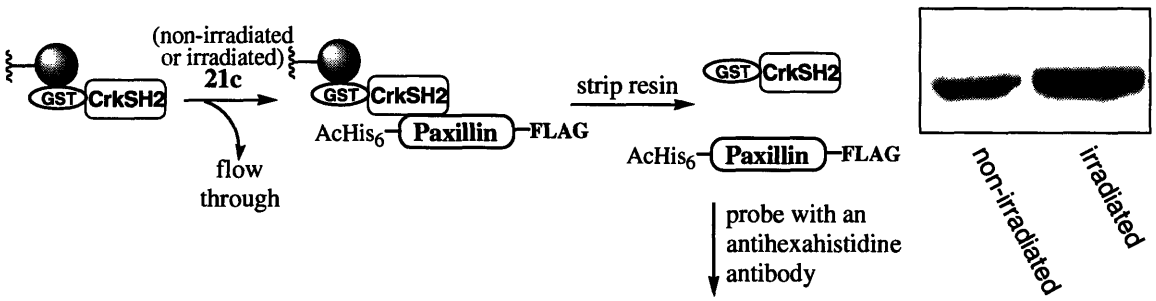


Figure 4.10. Western blot analysis of an in vitro binding assay between the Crk SH2 domain and cpY31 paxillin (**21c**) prior to and following irradiation. GST-CrkSH2 was bound to GST resin, incubated with **21c** either prior to (lane 1) or following (lane 2) irradiation for 90 s with 365 nm light. The resin was washed copiously and probed with an anti-hexahistidine primary antibody to visualize bound paxillin.

4-5. Focal adhesion turnover assays (with Dr. Steve Pratt, UVa)

Next, cpY31Pax (**21c**) was used to examine the role of Tyr31 paxillin phosphorylation on focal adhesion turnover in paxillin-null MEF cells. Preliminary experiments were carried out to assess whether phosphorylation at that site influenced adhesion dynamics. In collaboration with Professor Rick Horwitz' lab at UVa, we looked for a difference in the half life ($t_{1/2}$) of adhesion formation or disassembly in paxillin null cells microinjected with cpY31Pax (**21c**) prior to and following irradiation centered at $\lambda = 365$ nm. In a two week stay in the Horwitz lab, we were able to optimize a number of assay conditions, including the cell type, the choice of fluorescent marker, and the general protocol for cell plating and microinjection with the paxillin proteins. We then ran one to two trials of the focal adhesion assay each with **21b,** and caged and uncaged **21c**.

Focal adhesion turnover was observed using total internal reflection fluorescence (TIRF) microscopy to monitor GFP-labeled vinculin, which was transfected into cells, as a marker of focal adhesions. Vinculin is a structural protein that localizes to focal adhesions, and the use of fluorescently labeled vinculin has been reported for focal adhesion visualization.[16] The intensity of the focal adhesions over time was quantified using ImageJ software to determine a $t_{1/2}$ of assembly.

For these assays, we examined the focal adhesion dynamics of MEF cells injected with pY31Pax (**21b**), and with cpY31Pax (**21c**) prior to and following global irradiation.

Paxillin-null MEF cells were transfected with GFP-vinculin and plated on fibronectin-coated plates to provide a surface for the fibroblasts to spread. The plated cells were microinjected (typically 20 cells per plate) with **21b** or **21c** in a solution containing red dextran for visual identification of injected cells. For the uncaging experiment, the entire plate was irradiated with light centered at λ = 365 nm for 60 seconds on a standard DNA transilluminator following microinjection with **21c**. Cells containing both GFP and red dextran, indicating the presence of both the focal adhesion marker and microinjected protein, were selected for observation. Time-lapse fluorescence imaging was carried out for 15 to 30 minutes per cell on a TIRF microscope. Since MEF cells typically migrate for less than two hours after plating, only one to two cells from each plate were imaged. The focal adhesion assay is illustrated with **21c**-injected and uncaged cells in **Fig. 4.11**.

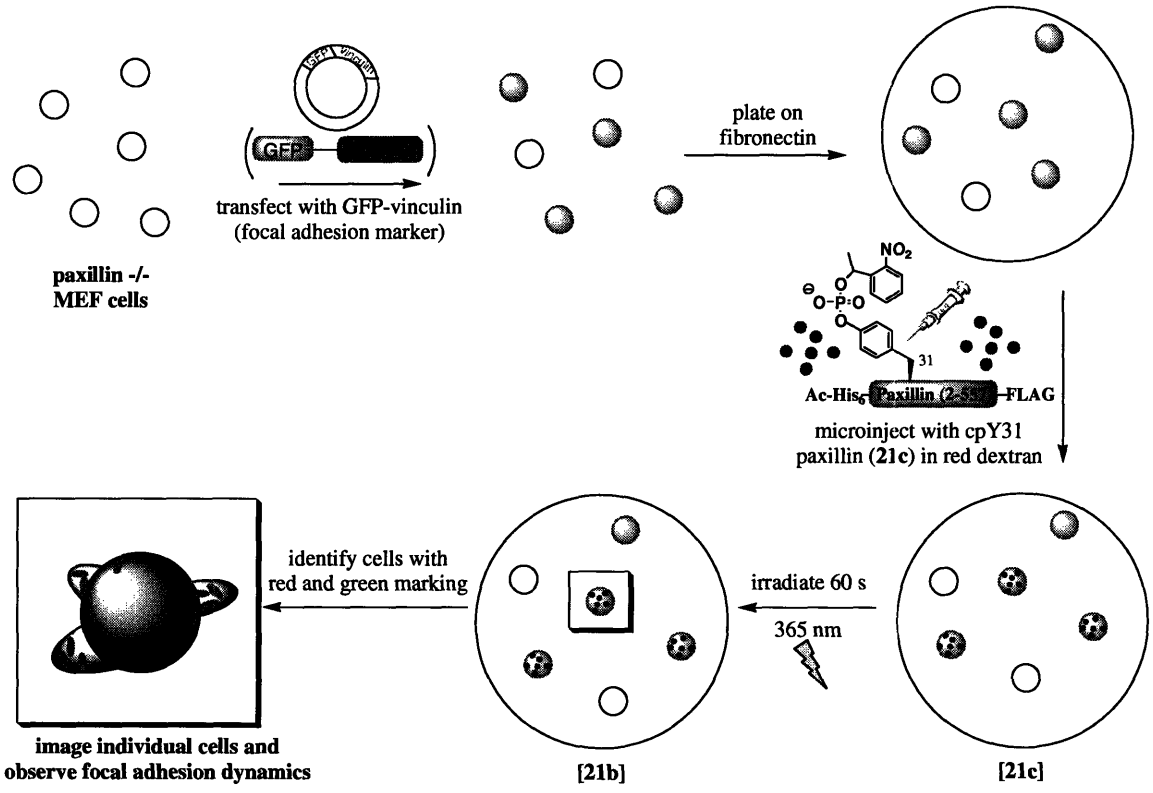


Figure 4.11. Assay protocol to probe the effect of Tyr31-paxillin phosphorylation on focal adhesion dynamics.

In cells microinjected with cpY31Pax (**21c**) and observed prior to uncaging, focal adhesions could be identified, but no new protrusions or focal adhesions formed, and there was no adhesion turn over. These observations were consistent with the phenotype of non-injected paxillin-null cells. However, after global uncaging, new protrusions formed, and focal adhesion formation and turnover was detectible in the protrusive areas. Similar adhesion turnover was visible in cells injected with pY31Pax (**21b**) (**Fig. 4.12a**).

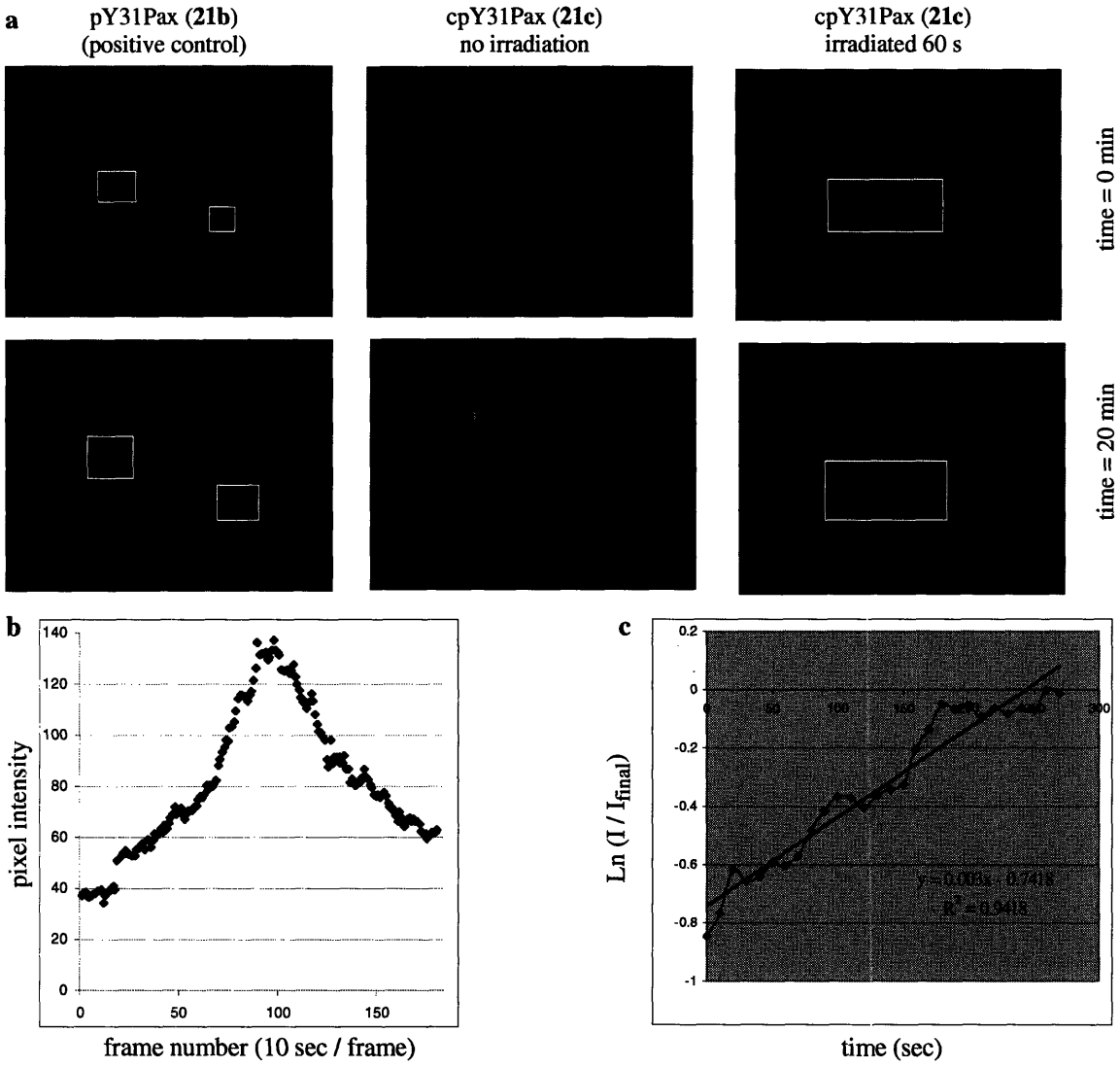


Figure 4.12. **a**) Snapshots from focal adhesion movies of MEF cells at time t= 0 and t = 20 minutes of imaging. The microscope imaged GFP-vinculin as a marker of focal adhesions. Boxes indicate protrusive areas with focal adhesion turnover. B) Graph of pixel intensity over time for a single focal adhesion with a pY31Pax (**21b**) microinjected cell. C) Graph of LN(I/Ifinal) over time, where I = pixel intensity for adhesion formation. The slope from this graph indicates a $t_{1/2}$ assembly = LN(2)/0.003 = 231 seconds (3.9 minutes).

The half life of focal adhesion formation was calculated for several adhesions in the imaged cells (**Fig. 4.12b,c**). While too few cells were imaged and adhesions analyzed to draw statistically significant conclusions, the preliminary data suggests a half life of focal adhesion formation of 3 to 5 minutes for the cells injected with pY31Pax (**21b**) or with cpY31Pax (**21c**) followed by uncaging. Confirmation of those results would suggest that phosphorylation at tyrosine 31 is essential for rescuing the adhesion turnover capacity of the cells.

Future experiments

Based on the promising results from our initial investigations, the following set of experiments and controls has been planned: 1) performing localized irradiation on individual cells, 2) imaging a statistically significant number of cells and adhesions for analysis, and 3) repeating the experiment with controls to standardize the influence of protein microinjection using expressed full-length paxillin, and Y31F and Y31E paxillin mutants.

The initial adhesion turnover experiments using cells injected with **21c** were carried out with global uncaging, meaning the entire plate of cells was irradiated prior to imaging. Future experiments, in contrast, will be performed with local uncaging, using a laser set up on the TIRF microscope to irradiate an individual cell or subcellular location. This will enable spatial and temporal information, which was not possible with the preliminary global uncaging experiments, by allowing comparison of the adhesion turnover in a single cell prior to and following irradiation and enabling comparison of the effect of a concentration burst of phosphoprotein between, for example, the leading edge versus the central region of a cell. In addition, local uncaging will provide a powerful internal control, since a specific cell can be imaged immediately preceding and immediately following **21c** uncaging. The chance of locating an identical cell prior to and following global irradiation is very low, since there are hundreds of cells per plate.

In collaboration with Dr. Steve Pratt in the Horwitz lab (UVa), the adhesion turnover assay will be repeated using laser-mediated uncaging with at least 5 separate cells to analyze a total of 15 to 30 adhesions. This number is based on related study of focal adhesion dynamics from the Horwitz lab,[16] and should provide sufficient data to

draw statistically significant conclusions. The semisynthetic strategy used to construct **21c** allowed access to milligram quantities of material, which are sufficient for hundreds of cell experiments.

Finally, the adhesion assay will be repeated using expressed paxillin controls, full-length paxillin, Y31F paxillin, and Y31E paxillin. Since the focal adhesion assay used here has previously been run exclusively with transfected cells, the controls will allow further investigation of any complications with the microinjection and provide more traditional biological controls. Mutants with a Tyr to Phe substitution are often used as phosphorylation-resistant protein mimics. Similarly, Ser to Asp or Ser to Glu substitutions are used to mimic constitutively phosphorylated Ser residues. The Y31E paxillin mutant was requested by Professor Horwitz as a possible phosphoTyr31 mimic, since there is no isosteric option from the 20 natural amino acids. Toward that end, two paxillin mutants were constructed, GST-ENLYFQ-Gly-Pax(1-557, Y31F)-FLAG, and GST-ENLYFQ-Gly-Pax(1-557, Y31F)-FLAG. The proteins were expressed, as described for **20** (Chapter 3) and treated with TEV protease to remove the 27-kDa GST tag. The resulting proteins, Y31F-FLAG and Y31E-FLAG will be used as described for **21b** in the focal adhesion assay.

The outlined experiments should provide sufficient data for dissecting the individual effect of Tyr31-paxillin phosphorylation on focal adhesion turnover in MEF cells, with specific temporal and spatial information.

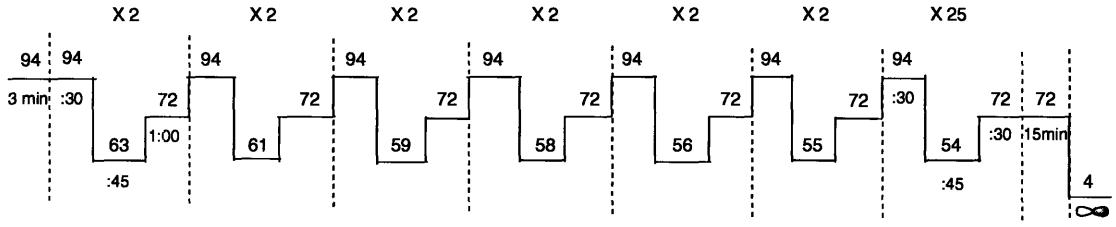
Conclusion

Binding and phosphorylation studies with **21a**, including FAK, GIT and PTP-PEST binding, and phosphorylation by Src, ERK, and JNK, confirm that the semisynthetic paxillin derivatives function comparably in vitro to native paxillin. Uncaging and CrkSH2 binding experiments with the semisynthetic caged phosphoprotein cpY31Pax (**21c**) verify that the probe releases phosphoTyr31 paxillin upon irradiation and creates a CrkSH2 binding platform for increased association with Crk. Therefore, the semisynthetic probe enables the isolated investigation of a single phosphorylation site in the context of the entire paxillin protein. Apart from residue 31, all phosphorylation and binding sites can participate in signaling and localization as with native paxillin.

Following the in vitro characterization described in this chapter, **21c** was used to investigate the influence of paxillin phosphorylation on focal adhesion turnover in paxillin-null MEF cells. Preliminary results from the adhesion turnover experiments suggest a crucial role for tyrosine 31 phosphorylation in focal adhesion dynamics. Ongoing experiments will provide time- and location-specific information on the downstream effects of tyrosine 31 phosphorylation. Further, the probe can be applied to additional in vivo assays to investigate other aspects of cellular migration.

Methods

Cloning of GIT(622-761) into pGEX-4T-2. A kan-resistant GIT construct (supplied by Rick Horwitz) was transformed into DH5α cells and grown up on an LB-kan agar plate. Several colonies were selected and grown up in 3 mL of LB-kan overnight with shaking at 37 °C. The plasmids from these cultures were isolated using a perfect prep plasmid mini kit (Eppendorf) following the manufacturer's instructions. GIT(622-761) was amplified from the plasmid DNA from a single miniprep using primers to introduce a 5' EcoR1 site and a 3' Not1 site for incorporation into a pGEX-4T-2 vector, which provides an *N*-terminal glutathione S-transferase (GST) moiety. For this amplification the following primers were used: *FORGIT121804*: 5'- GCC GGA ATT CGT GAA GAC TTC CAC CCA GAG CTG GAA AGC -3' and *BACGIT121804*: 5'- GCC CCC TTT TGC GGC CGC TCA CTG CTT CTT CTC TCG GGT GGT GAT GGT-3' (Invitrogen). Five 50-μL PCR reactions were run, each with 40 μL sterile DI water, 5 μL 10X Taq HIFI buffer, 0.35 μL 100 mM dNTPs, 2 μL 50 mM $MgCl_2$, 0.65 μg each of FORGIT121804 and BACGIT121804, 0.7 μL of GIT miniprep A DNA template, and 0.5 μL of Taq HIFI (added last). The reactions were run in the thermal cycler as depicted below:



The purified GIT insert and a pGEX-4T-2 vector were digested with EcoR1 and Not1. To 43.5 μL of purified PCR product in EB were added 5.0 μL 10X EcoR1 buffer,

0.5 μL 100X BSA, and 2.5 μL Not1. The digestion was incubated overnight at rt, 2.5 μL of EcoR1 were added, and the mixture was further incubated for 1 hour at 37 °C. To this solution was then added 12.5 μL 5X nucleic acid loading buffer, and the digested insert was electrophoresed on a 1.5% agarose DNA gel for purification. For the pGEX-4T-2 vector digestion, 750 ng of vector (10 μL) were combined with 5.0 μL 10X EcoR1 buffer, 0.5 μL 100X BSA, 29.5 μL sterile water, and 2.5 μL Not1, and the mixture was incubated overnight at rt. EcoR1 (2.5 μL) was added, and the mixture was incubated as described for the purified PCR product. The enzymes were deactivated by heating for 20 min at 65 °C, and 15 μL of CIP was subsequently added to the vector digestion and incubated for 1 hour at 37 °C. To the solution was then added 12.5 μL of 5X nucleic acid loading buffer, and the cut/CIP-treated vector was electrophoresed on a 1.5% agarose gel. The vector and insert were excised and purified with a gel extraction kit (Qiagen). The DNA was eluted with 30 μL of sterile water and used directly in ligations without further concentration. DNA not immediately ligated was stored at -20 °C.

Ligation of the GIT(622-761) insert into the pGEX-4T-2 vector was accomplished by combining 3 μL vector, 5 μL PCR insert, 1 μL T4 ligase, and 1 μL 10X T4 ligase buffer in a sterile eppendorf. The tube was gently flicked, then incubated overnight (although 2-3 h should be sufficient) at room temperature. A vector-only control ligation was also performed using 3 μL vector, 1 μL T4 ligase, 1 μL 10X T4 ligase buffer, and 5 μL sterile water. The mixtures were placed on ice and used immediately for transformation into MAX Efficiency DH5α chemically competent cells.

Cloning of FAK(857-1053) into pGEX-4T-2. An amp-resistant FAK construct (donated by Tom Parson) was transformed into DH5α cells and grown up on an LB-carbenicillin agar plate. Several colonies were selected and grown up in 3 mL of LB-carbenicillin overnight with shaking at 37 °C. The plasmids from these cultures were isolated using a perfect prep plasmid mini kit (Eppendorf). FAK(857-1053) was amplified from plasmid DNA with PCR using the following primers to introduce a 5' EcoR1 site and a 3' NOT1 site for incorporation into a pGEX-4T-2 vector: *FORFAK121804*: 5'- GCC GGA ATT CGT GCT GGT AAC CAG CAC ATA TAT CAG CCT -3' and *BACFAK121804*: 5'- GCC CCC TTT TGC GGC CGC TTA GTG GGG CCT GGA CTG GCT GAT CAT

TTT-3' (Invitrogen). The PCR-amplified fragment was digested with EcoR1 and Not1 and ligated to the EcoR1/Not1 digested and CIP-treated pGEX-4T-2 vector as described above with GIT.

Transformation of GIT and FAK constructs into DH5α cells. A 90-μL aliquot of MAX Efficiency DH5α chemically competent cells was thawed on ice. The cells were mixed gently with a pipette, and 30-μL aliquots were transferred into each of 3 pre-chilled Falcon tubes. As a positive control, 1 μL (10 pg) of supplied pUC19 DNA was added to one of the Falcon tubes and carefully mixed by flicking. To the remaining two Falcon tubes were added 1 μL of the GST-GIT(622-761) or GST-FAK(857-1053) ligation mixtures. After gentle tapping, the cells were incubated on ice for 30 min, heat shocked in a 42 °C water bath for exactly 45 seconds and placed on ice for 2 min. Room temperature SOC media (0.27 mL) was added to each tube, and the tubes were incubated at 37 °C for one hour. 100 μL of each sample were spread onto separate LB-carbenicillin agar plates along with enough SOC media to bring the final volume to 100 μL. The plates were incubated at 37 °C overnight and stored at 4 °C. Hundreds of colonies grew up on the positive control, on the FAK, and 30 on the GIT vector/insert ligation plates.

To screen the transformants, five colonies were selected from each plate and grown up in 3 mL of LB-carbenicillin overnight with shaking at 37 °C. The plasmids from these cultures were isolated using plasmid mini kit. The isolated plasmids were then screened for the presence of GIT(622-761) or FAK(857-1053) with Not/EcoR1 digestion to excise the 423 and 594 bp insertions. The constructs were then confirmed by DNA sequencing. For each submission, the following was added to a 600-μL eppendorf tube: 0.62 μL (3.2 pmol) of 5' pGEX primer, 6 μL (estimated between 200-500 ng) of a GIT or FAK miniprepped DNA, and 5.36 μL sterile water.

Preparation of GST-PTP-PEST(338-390) for transformation and expression. A stab culture of pGEX PTP-PEST(338-390) in amp-resistant DH5α cells (donated by Mike Schaller) was grown up in LB with carbenicillin and then reselected on an

LB/carbenicillin plate at 37°C for 16 h. Hundreds of colonies grew up, and two were selected and miniprepped for subsequent transformation into an expression cell line.

Expression and GST-affinity purification of GST, GST-GIT(622-761), GST-FAK(857-1057), and GST-PTP-PEST(338-390). The isolated plasmids from an empty pGEX-4T-2 vector, and from GIT, FAK and PTP-PEST minipreps were transformed into BL21-CodonPlus®-RP or RIL competent cells (Stratagene) and grown to midlog phase in 1 L of LB media with carbenicillin and chloramphenicol at 37 °C. The cells were induced with 0.1 mM IPTG, and protein expression proceeded for 4 hours at 37 °C. Cells were harvested by centrifugation, and the pellets (ranging from 6 to 9 g) were stored at –80 °C. For cell lysis, pellets were thawed and resuspended in 35 mL of lysis buffer (PBS, 1 mg/mL lysozyme, 1 mM DTT, and protease cocktail III [100 μM AEBSF, 80 nM aprotinin, 5 μM bestatin, 1.5 μM E-64, 2 μM leupeptin, 1 μM pepstatin A] (Calbiochem))and incubated for 20 min. at 4 °C. The cells were lysed with 1% NP-40 alternative, then sonicated and subjected to centrifugation at 13,000 rpm for 30 min, followed by 35,000 rpm for 30 min or by filtration through a 0.2 micron filter.

The soluble fraction for each protein was purified using Glutathione Sepharose 4 Fast Flow resin (Amersham Biosciences). Each protein solution was incubated with 3 mL of resin at 4 °C for 1 hour. The resin was washed with cold PBS (150 mL), and four 3-mL elutions were collected. For each elution, the resin was incubated with 10 mM glutathione in 50 mM Tris, pH 8.0 for 3-5 min at rt. Elutions for each protein were combined and dialyzed into PBS for subsequent binding studies. The proteins were analyzed by 10% SDS-PAGE gels and visualized with Coomassie blue dye. The purified protein was stored at 4 °C and used within three weeks of purification.

GST pull-downs of AcHis$_6$-Pax(2-557)-FLAG, with FAK, GIT, PTP-PEST, and GST. Glutathione sepharose beads were loaded with GST-FAK(857-1053), GST-GIT(622-761), GST-PTP-PEST(338-390), and GST by incubating 50 μL of resin (maximum binding capacity = 500 μg) in each of 4 1.5-mL eppendorf tubes with 1 mg of protein (500 μL of 2 mg/mL solutions) in PBS for 30 min. The tubes were sedimented by centrifugation at 1000 g, and the supernatant was discarded. The resin was washed with

3 x 500 μL of PBS. To each tube was added semi-synthetic Y31Pax (~ 200 μg) in 600 μL of PBS with 1% BSA. The mixture was agitated for 1 hour, and the glutathione sepharose beads were sedimented by centrifugation. The resin was washed with PBS (3 x 500 μL) and boiled in 2X SDS reducing gel loading buffer (100 μL). The samples were analyzed by 10% SDS-PAGE gels, and visualized by Western blot analysis with a mouse anti-hexahistidine primary antibody and a goat anti-rabbit alkaline phosphatase secondary antibody.

*Src, ERK, and JNK phosphorylation of expressed and semisynthetic paxillin (**21c**).* For the kinase assays, all kinases were stored in 1 to 4 μL aliquots at -80 °C. The kinases (Biosource) were used as described below: Src (PHO3114), ERK2(PHO3104), JNK2α2(PHO3014), and JNK1α1 (Upstate). For long term storage 2.66 M $MgCl_2$ was kept at room temperature in a 0.1 M HCl solution to prevent crashing out. Prior to use, a 1/10 dilution was prepared in TBS.

To 2 μg of paxillin was added 8 μL of 1 mM ATP (200 μM final concentration), 2.25 μL of 226 mM $MgCl_2$ (15 mM final concentration), and 1.5 μL c-Src (450 units), or 4 μL ERK2 (148 units), or 2 μL of JNK (128 units). TBS, pH 7.5, was added to a final volume of 40 μL, and the contents were mixed by pipetting seven times and gently flicking. The reaction was incubated for 15 min at 30 °C and quenched with the addition of 40 μL of 2X SDS reducing gel loading buffer to give a final substrate concentration of 25 ng/μL. A 1/5 dilution of this solution was prepared for the more sensitive Western blots. For detection by pan and pY-specific antibodies, 20 to 30 ng (4 to 6 μL of the 1/5 dilution) were loaded per well. For detection by pSer-specific antibodies, 250 ng of substrate (10 μL) were loaded per well. Gels were run on a 10% PAGE at 140V for about 2 hours before being transferred to nitrocellulose.

Western blotting for chemiluminescent detection. The following primary and secondary antibodies (Biosource) were used to probe for various sites of paxillin phosphorylation: mouse (monoclonal) anti-human paxillin (AHO0492), rabbit (polyclonal) anti-paxillin[pY31] (44-720G), rabbit (polyclonal) anti-paxillin[pY118] (Biosource), rabbit (polyclonal) anti-paxillin[pS126] (44-1022G), rabbit (polyclonal) anti-paxillin[pS178]

(44-1026G), goat anti-rabbit HRP conjugate (ALI3403), and goat anti-mouse HRP conjugate (AMI4404).

Each Western blot nitrocellulose membrane was blocked for 1.5 to 2 hours at RT with Superblock (Pierce) with 0.05% Tween-20 and washed three times with 20 mL of TBST for 5 min. each. The membrane was then incubated with a pan (1/1,000 dilution) or phospho-specific (1/2,000 dilution) paxillin primary antibody in TBST for 1 hour at RT. The membrane was rinsed with 5 x 5 minute washes of TBST. A goat anti-mouse (1/10,00 dilution) or goat anti-rabbit (1/20,000 dilution) HRP-conjugated secondary antibody in TBST was incubated with the membrane for 1 hour at RT. The membrane was rinsed again with 5 x 5 min. washes of TBST. Supersignal west chemiluminescent substrate (Pierce, 34080) was added in a sufficient amount to fully cover the block (about 5 mL per blot) and incubated for 3 to 5 min. The chemiluminescent substrate was decanted, and the corner of the membrane was touched onto a paper towel to remove as much of the reagent as possible. The membrane was placed into a clear plastic pocket and exposed for 3 to 5 minutes.

Uncaging calculations. For uncaging quantification, a 1.2 mg/mL solution of cpY31Pax (**21c**) in TBS with 2.5 mM DTT was irradiated for 90 s in a glass cell (pathlegnth 1 mm) at 365 nm with an intensity of 7,330 $\mu W/cm^2$ on a DNA transilluminator. Semisynthetic pY31Pax (**21b**) and equal amounts of caged and uncaged protein were run on 10% SDS gels and transferred to nitrocellulose. Western blots were developed with mouse (monoclonal) anti-human paxillin and phosphospecific rabbit (polyclonal) anti-paxillin[pY31] primary antibodies and visualized by chemiluminescence. Pixel counting was used to calculate the relative amount, compared to a pY31Pax (**21b**) standard, of phosphoprotein at t = 0 and t = 90 seconds after uncaging of cpY31Pax (**21c**). A sample calculation is shown below with data from the Western blot shown in **Fig. 4.9c**.

$$\% \text{ uncaged} = \frac{\text{normalized intensity of total pY31paxillin detected}}{\text{normalized intensity of total paxillin detected}}$$

% of pY31Pax uncaged at t = 0 seconds = 0.040/0.619 = 0.06 = **6%**

% of pY31Pax uncaged at t = 90 seconds = 0.524/0.599 = 0.87 = **87%**

Data table for the anti-paxillin (pan) probed Western blot

	Pixel count	**Pixel count - blank**	**normalized intensity**
Blank (B1)	2211	-	-
pY31Pax (U3)	7711	5500	**1**
cpY31Pax, t = 90 (U2)	5508	3297	**0.599**
cpY31Pax, t = 0 (U1)	5615	3404	**0.619**

Data table for the anti-[pY31]paxillin (phosphospecific) probed Western blot

	Pixel count	Pixel count - blank	normalized brightness
Blank (B1)	2272	-	-
pY31Pax (U3)	10650	8378	**1**
cpY31Pax, t = 90 (U2)	6660	4388	**0.524**
cpY31Pax, t = 0 (U1)	2609	337	**0.040**

Crk-Binding Assay. Glutathione sepharose beads were loaded with GST-CrkSH2 by incubating 40 μL of resin (maximum binding capacity = 400 μg) in two 1.5-mL eppendorf tubes with 500 μg each of GST-CrkSH2 (1.2 mL) in PBS for 1 hour. The supernatant was discarded, and the resin was washed with PBS (3 x 1.2 mL). Meanwhile, 35 μg of cpY31Pax (**21c**) in PBS (100 μL) with 5 mM DTT was irradiated with light centered at 365 nm in a 1 mm glass cell on a DNA transilluminator for 90 s. The solution was diluted to a final volume of 1.2 mL PBS with 1% BSA. A second sample of cpY31Pax was prepared identically but was spared the UV irradiation. The GST-CrkSH2-loaded resin was incubated with either the caged or uncaged **21c** solution for 2 h. The paxillin solution was removed, and the resin was washed with PBS (2 x 1.3 mL) prior to being boiled in 2X SDS reducing gel loading buffer (100 μL). The samples

were analyzed by 10% SDS-PAGE gels and visualized by Western blot analysis probed with a mouse anti-hexahistidine primary antibody and a goat anti-rabbit alkaline phosphatase secondary antibody.

Transfection with GFP-vinculin. Paxillin-null MEF cells were provided by the Horwitz lab (University of Virginia). For transfection, a GFP-vinculin construct (1.5 μL of a 0.7 mg/mL solution) was re-suspended by tapping and added to a 1.5-mL eppendorf tube. Pre-warmed Opti-MEM-1 reduced serum media (100 μL), and liptofectamine (5 μL) were added, and the mixture was flicked to mix. Following incubation at rt for 30 to 60 min, additional Opti-MEM-1 reduced serum media (900 μL) was added. Paxillin-null MEF cells were removed from incubation at 37 °C and washed twice with Opti-MEM-1 reduced serum media. The media was removed from the cell plates, and 1 mL of the DNA mixture was slowly added drop-by-drop over the MEFs. The cells were incubated for 3 hours at 37 °C and washed with complete media. The cells were then incubated overnight in complete media (3 mL) and used within 24 to 36 hours.

Focal adhesion dynamics assay. Paxillin-null MEF cells transfected with GFP-vinculin were plated on fibronectin-coated tissue culture dishes. In brief, the cells were washed with PBS, and trypsin/EDTA (0.75 mL) was added to the cells and incubated for 5 min at 37 °C. Meanwhile, Opti-MEM complete media (2 mL) was added to a fibronectin-coated (1 μg/mL) glass-bottomed plate. Complete media (4 mL) was added to the cells, and 1/5 of the total volume of cells in media was added to the fibronectin-coated plate. The plated cells were incubated at 37 °C for 20 to minutes, and then 20 to 30 cells per plate were microinjected with pY31Pax (**21b**) or cpY31Pax (**21c**) (0.5 mg/mL) in PBS with red dextran. All microinjection was performed by Brian Hall in the (Schwartz lab, University of Virginia). Microinjection typically lasted 20 to 30 min. The cell media was exchanged with fresh Opti-MEM complete media, and the cells were either observed immediately or first irradiated for 60 sec on a standard DNA transilluminator (7,330 $\mu W/cm^2$) with light centered at λ = 365 nm. Cells visible via GFP and red dextran imaging were identified, and one to two cells per plate were observed using time lapse fluorescence imaging (10 sec/frame) for 30 min using a TIRF microscope (Olympus).

The live cell imaging was performed by Steve Pratt (Horwitz lab, Universtiy of Virginia). Analysis was carried out using ImageJ software to quantify the fluorescence intensity of individual focal adhesions, using GFP-labeled vinculin as a marker for focal adhesions, over time. The semilogarithmic plots of the background-subtracted fluorescence intensity over time was linear, and the half life of adhesion formation was determined from the resulting slopes.[3]

GST-paxillin(1-557)-FLAG, Y31E, and Y31F paxillin mutants. The full-length paxillin construct and the Y31E and Y31F mutants were expressed and purified by GST and FLAG-affinity purification as described for GST-paxillin(38-557)-FLAG (**21**) in Chapter 3. TEV protease cleavage, performed as described in Chapter 3, yielded native, Y31E, and Y31F Gly-paxillin(1-557)-FLAG. The proteins were stored at 1 mg/mL in PBS with 20% glycerol at -20 °C.

Acknowledgements: I thank Professor Rick Horwitz for welcoming me into his lab to perform the focal adhesion experiments, and I thank Steve Pratt and Anjana Nayal for their contribution to the focal adhesion studies. Steve ran the TIRF microscope and Anjana taught me the necessary cell culture and transfection procedures. I thank Brian Hall for skillfully performing all the microinjections. I thank Professor Rick Horwitz, Professor Tom Parsons, and Professor Mike Schaller for the generous donation of GIT, FAK, and PTP-PEST constructs, respectively.

References

1. Lo, S. H. Focal adhesions: What's new inside. *Dev. Biol.* **294**, 280-291 (2006).
2. Jockusch, B. M. et al. The molecular architecture of focal adhesions. *Annu. Rev. Cell Dev. Biol.* **11**, 379-416 (1995).
3. Webb, D. J. et al. FAK-Src signaling through paxillin, ERK and MLCK regulates adhesion disassembly. *Nat. Cell Biol.* **6**, 154-161 (2004).

4. Brown, M. C. & Turner, C. E. Paxillin: Adapting to change. *Physiol. Rev.* **84**, 1315-1339 (2004).

5. Hagel, M. et al. The adaptor protein paxillin is essential for normal development in the mouse and is a critical transducer of fibronectin signaling. *Mol. Cell. Biol.* **22**, 901-915 (2002).

6. Hildebrand, J. D., Schaller, M. D. & Parsons, J. T. Paxillin, a tyrosine phosphorylated focal adhesion-associated protein binds to the carboxyl terminal domain of focal adhesion kinase. *Mol. Biol. Cell* **6**, 637-47 (1995).

7. Manabe, R.-I., Kovalenko, M., Webb, D. J. & Horwitz, A. R. GIT1 functions in a motile, multi-molecular signaling complex that regulates protrusive activity and cell migration. *J. Cell Sci.* **115**, 1497-1510 (2002).

8. Shen, Y., Schneider, G., Cloutier, J.-F., Veillette, A. & Schaller, M. D. Direct association of protein-tyrosine phosphatase PTP-PEST with paxillin. *J. Biol. Chem.* **273**, 6474-6481 (1998).

9. Huang, C., Rajfur, Z., Borchers, C., Schaller, M. D. & Jacobson, K. JNK phosphorylates paxillin and regulates cell migration. *Nature (London, United Kingdom)* **424**, 219-223 (2003).

10. Rothman, D. M., Vazquez, M. E., Vogel, E. M. & Imperiali, B. General Method for the Synthesis of Caged Phosphopeptides: Tools for the Exploration of Signal Transduction Pathways. *Org. Lett.* **4**, 2865-2868 (2002).

11. Humphrey, D. et al. In Situ Photoactivation of a Caged Phosphotyrosine Peptide Derived from Focal Adhesion Kinase Temporarily Halts Lamellar Extension of Single Migrating Tumor Cells. *J. Biol. Chem.* **280**, 22091-22101 (2005).

12. Nguyen, A., Rothman, D. M., Stehn, J., Imperiali, B. & Yaffe, M. B. Caged phosphopeptides reveal a temporal role for 14-3-3 in G1 arrest and S-phase checkpoint function. *Nat. Biotechnol.* **22**, 993-1000 (2004).

13. Kaplan, J. H., Forbush, B., 3rd & Hoffman, J. F. Rapid photolytic release of adenosine 5'-triphosphate from a protected analogue: utilization by the Na:K pump of human red blood cell ghosts. *Biochemistry* **17**, 1929-35 (1978).

14. Birge, R. B. et al. Identification and characterization of a high-affinity interaction between v-Crk and tyrosine-phosphorylated paxillin in CT10-transformed fibroblasts. *Mol. Cell. Biol.* **13**, 4648-56 (1993).

15. Schaller, M. D. & Parsons, J. T. pp125FAK-dependent tyrosine phosphorylation of paxillin creates a high-affinity binding site for Crk. *Mol. Cell. Biol.* **15**, 2635-45 (1995).

16. Nayal, A. et al. Paxillin phosphorylation at Ser273 localizes a GIT1-PIX-PAK complex and regulates adhesion and protrusion dynamics. *J. Cell Biol.* **173**, 587-599 (2006).

Chapter 5

Semisynthesis of a Sox-based chemosensor for ERK2 phosphorylation

Introduction

Sensors for kinase-mediated phosphorylation enable the investigation of kinase activities in signal transduction pathways. Aberrant kinase activity is known to be involved in a number of human diseases, including many forms of cancer.[1] Therefore, kinase-activity sensors can be valuable for screening potential kinase inhibitors for therapeutic applications.[2] Furthermore, kinase sensors provide a tool to help dissect the role of individual molecules in complex phosphorylation-based signaling cascades.

Sox-based fluorescent chemosensors of protein kinase activity

Sox-based chemosensors are short peptides sensors that include a kinase recognition motif and a chelation-sensitive fluorophore, the 8-hydroxyquinoline moiety, incorporated into the non-natural amino acid Sox (**Fig. 5.1a**).[3] The sensors rely on the Mg^{2+}-chelating property of the phosphate group to sense phosphorylation within an optimized target peptide sequence. Prior to phosphorylation, the Mg^{2+} binding affinity of the sensors is weak. Upon phosphorylation, however, the Mg^{2+} binding affinity significantly increases, and chelation between the Mg^{2+} and Sox results in chelation-enhanced fluorescence (CHEF) (**Fig. 5.1b**).[4] The optimal position of the Sox fluorophore relative to the phosphorylated residue for Mg^{2+} chelation is set by a β-turn designed into the linear peptide or by conjugating the Sox chromophore on the side chain of an appropriately-spaced cysteine residue. When the Sox chromophore is appended to a cysteine side chain, both *N*- and *C*-terminal kinase recognition elements can be incorporated into the sensor, which can lead to improved sensitivity and specificity. Sox-based sensors have a number of important properties that make them ideal for assaying kinase activity; they demonstrate a large fluorescence increase upon phosphorylation (typically three- to eight-fold),[2] they enable continuous fluorescence monitoring, and they can be applied to serine, threonine, or tyrosine kinases.[4] Sox-based peptide chemosensors have been developed for a number of important kinases in signal

transduction pathways, and the versatile sensor design is applicable to any kinase that phosphorylates a distinct linear peptide recognition motif.[2,4,5]

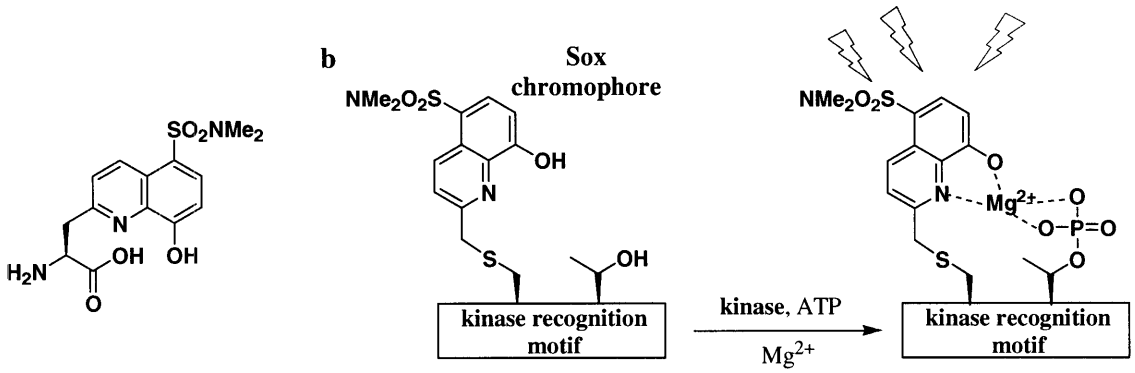


Figure 5.1. **a)** The Sox amino acid, which includes the 8-hydroxyquinoline chelation-sensitive fluorophore. **b)** Sox-based fluorescent chemosensors for kinase activity, shown here with a threonine as the phosphorylated residue.

Sox-based sensing of ERK2 activity

Two significant challenges in the design of phosphorylation sensors are achieving high specificity for detecting the activity of a single protein kinase and designing sensitive probes for kinases that have minimal recognition sequences. An individual kinase can have tens or hundreds of substrates.[6] Similarly, many phosphorylation sites can be phosphorylated with varying efficiency by a number of different kinases or kinase isoforms. The major recognition element for most kinases is the phosphorylation motif of the target substrate, a linear peptide sequence that is typically composed of eight or fewer residues. Intrinsically, a phosphorylation sensor can be only as specific or sensitive as the given kinase is for the linear target sequence on which the sensor is modeled.

In addition to recognizing a linear phosphorylation sequence, some kinases, particularly those that phosphorylate substrates with shorter or less unique consensus motifs, rely on recognition of extended regions, resulting in protein-protein interactions that extend beyond the active site. The mitogen-activated protein kinase (MAPK) ERK2 phosphorylates the transcription factor Ets-1 at Thr38 in the short T-P consensus motif.[7] ERK2 has been shown to interact with the pointed (PNT) domain of Ets-1, a five-helix bundle spanning residues 54-135 of the protein and flanking the unstructured *N*-terminus

(**Fig. 5.2a**).[8] An Ets-1 construct composed of residues 1-138 (EtsΔ138) was recently investigated by Dalby and coworkers in a study that demonstrates the essentiality of the PNT domain for optimal ERK2 phosphorylation.[9]

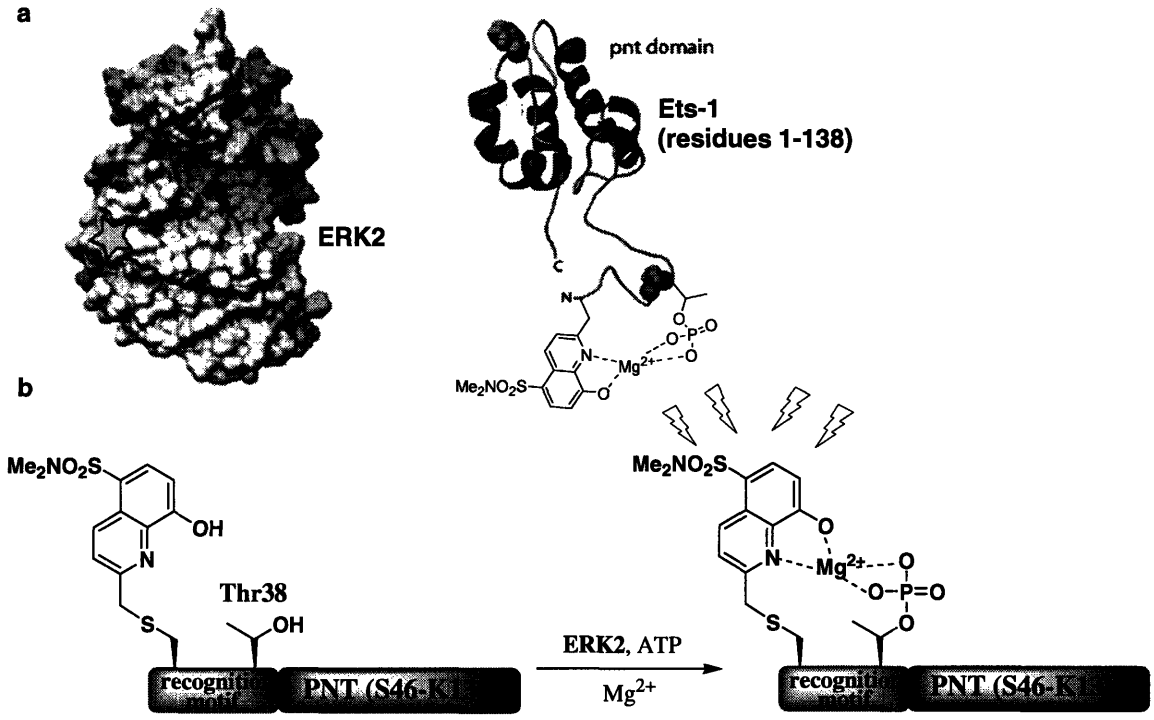


Figure 5.2. **a**) (Figure modified from Rainey and coworkers.)[9] Structure of the kinase ERK2 with the active site aspartic acid shown in red and the PNT docking site indicated with a yellow star. The structure of Ets-1 residues 1-138, including the PNT domain, which interacts with ERK2, and the TP phosphorylation motif, shown with the Sox amino acid. **b**) The Sox-based ERK2 kinase sensor. The sensor includes the PNT domain of Ets-1 and an ERK2 recognition motif that contains the phosphorylated residue (Thr38) and the Sox amino acid.

Previous attempts at designing a Sox-based sensor for ERK2 phosphorylation using linear peptide sequences have failed to achieve acceptable specificity, which could be explained by the more complex recognition strategy of ERK2. In this chapter, we describe the design and semisynthesis of a Sox-based ERK2 sensor that incorporates the entire PNT domain in addition to an ERK2 recognition motif (**Fig. 5.2b**). The contribution of the extended recognition element, the PNT domain, should confer the ERK2 sensor with excellent specificity. The sensor was constructed using NCL to conjugate the TP recognition motif for ERK2 and the Sox sensing module onto a

biologically expressed fragment comprising the Ets-1 PNT domain. The ligation design elements developed with the paxillin system (Chapter 3) were fully transferable to the semisynthesis of the ERK2 sensor.

Results and Discussion

5-1. Design of the Sox-based ERK2 phosphorylation sensor

The design for the ERK2 phosphorylation sensor (**22a**) incorporates a Sox-containing recognition peptide fused to a biologically-expressed fragment comprising residues 46 through 138 of Ets-1 and including the PNT domain. An optimized phosphorylation motif (V-P-C(Sox)-L-**T**-**P**-G-G-R-R) was used in place of the phosphorylated region of Ets-1 (L-L-**T**-**P**-S-S-K-E) in the sensor, and the native distance between the recognition sequence, T-P, and the PNT domain was preserved (**Fig. 5.3a**). The residue replacements occur in the unstructured *N*-terminal region of Ets-1, so secondary structures within the protein would not be predicted to be disrupted by the semisynthesis. Construction of the sensor by NCL required a *C*-terminal thioester on the *N*-terminal fragment and an *N*-terminal cysteine residue on the *C*-terminal fragment.[10] Accordingly, the ligation joined a synthetic Sox-containing peptide thioester, either Ac-VPC(**Sox**)LTPGGRRG-SBn (**PNTD-S-SBn**, **23a**) or Ac-VPC(**Sox**)L**pT**PGGRRG-SBn (**PNTD-P-SBn**, **23b**), to the expressed PNT domain. The design and construction of the ERK2 sensor (**22a**) and the phosphorylated ERK2 sensor (**22b**) are shown in **Figure 5.3**. A Gly-Cys ligation site was included to eliminate the possibility of epimerization at the *C*-terminal residue in the peptide synthesis and to increase ligation efficiency, as has been shown when Gly is the terminal thioester residue for NCL.[11]

For the *C*-terminal fragment of the sensor, residues 46-138 of Ets-1 were expressed as a GST-fusion protein with a hexahistidine tag included at the *C*-terminus to provide two handles for purification. A recognition peptide, ENLYFQ(✂)C, for TEV protease was positioned immediately *N*-terminal of the insert. Treatment of GST-ENLYFQC-PNT(46-138)-His_6 (**24**) with TEV protease resulted in the loss of the GST tag to expose an *N*-terminal cysteine residue and produce Cys-PNT(46-138)-His_6 (**25**). Ligation of the *C*-terminal fragment **25** with the synthesized Sox-containing thioesters yielded the ERK2 sensor (**22a**) and the corresponding phosphorylated sensor (**23b**). The

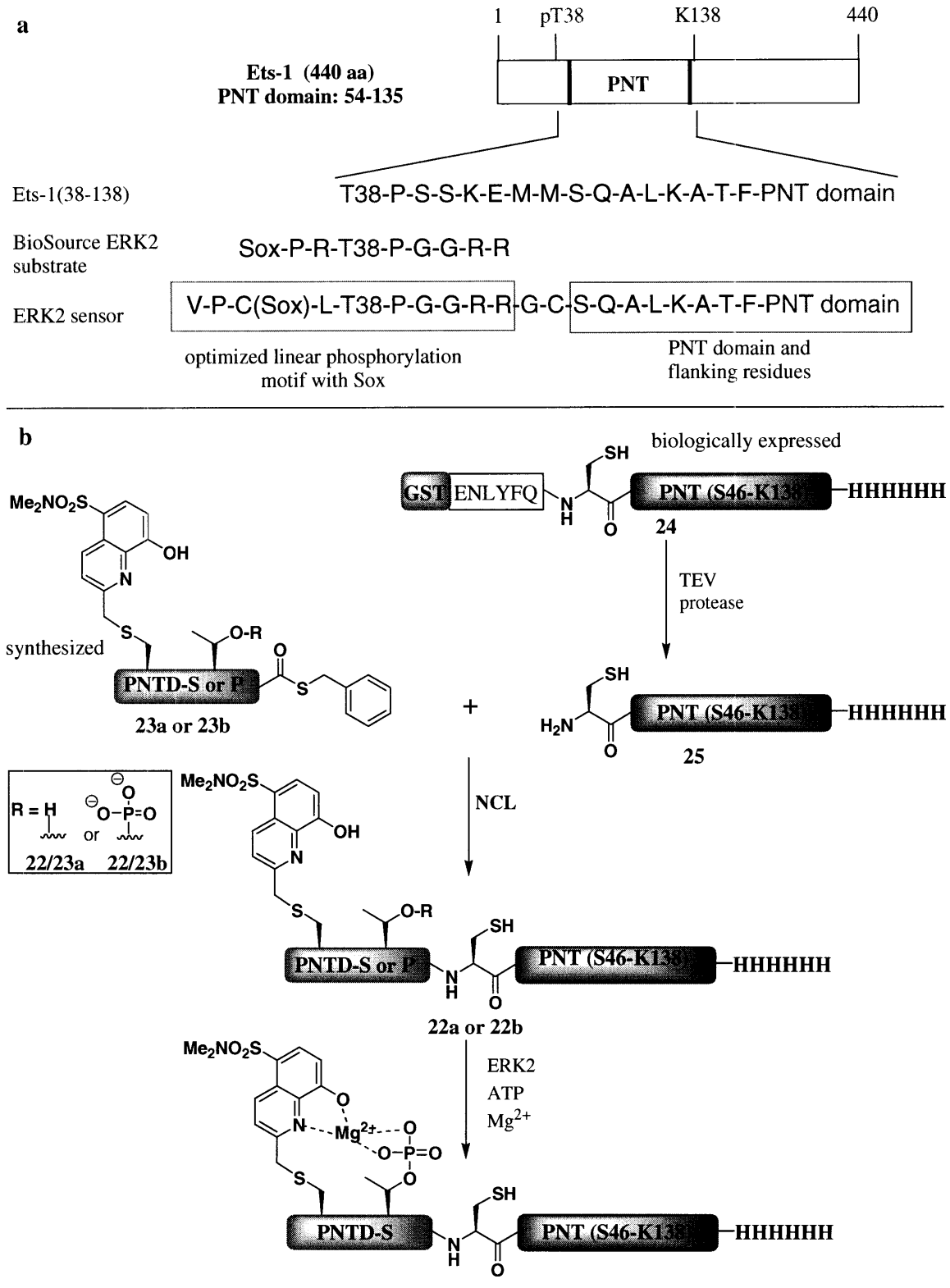


Figure 5.3. **a**) ERK2 Sox-sensor design. The sensor includes a linear recognition motif for ERK2 and the PNT domain of Ets-1. **b**) Semisynthesis of the ERK2 sensor (**22a**) and phosphorylated sensor (**22b**).

phosphorylated sensor (**22b**) was constructed in parallel to determine the fluorescence difference of **22a** prior to and following threonine phosphorylation for sensor calibration.

5-2. Preparation of the C-terminal sensor fragment (25)

PNT(46-138) was cloned from a kanamycin-resistant pET28a vector encoding Ets S26A. The PNT(46-138) insert was prepared for cloning into a pGEX-4T-2 GST-fusion vector using primers to create a hexahistidine tag at the *C*-terminus and a TEV protease site immediately *N*-terminal to the expressed PNT(46-138) fragment. Following PCR and ligation into the pGEX vector, the isolated plasmid DNA was confirmed by DNA sequencing to contain the correct insert. The DNA was transformed into BL21(DE3) competent cells for expression of GST-ENLYFQ-Cys-PNT(46-138)-His_6 (**24**). Protein expression was induced with 0.2 mM IPTG in LB-carbenicillin at 30 °C and carried out for 4 to 5 hours. Crude **24** was purified on a GST-affinity column and eluted from the column with 10 mM reduced glutathione. Gel analysis revealed a single species corresponding to the 38.5 kDa construct. As quantified using a BioRad protein assay, which is compatible with the presence of glutathione, a 1 L expression of **24** yielded 25 to 30 mg of GST-purified protein. Complete cleavage of the GST tag, revealing an *N*-terminal cysteine residue, was accomplished by treatment with TEV protease. By gel analysis, incubation of **24** with TEV protease for 3 hours resulted in the disappearance of the 38.5 kDa species and the concomitant appearance of an 11.5 kDa fragment, consistent with the calculated molecular weight for Cys-PNT(46-138)-His_6 (**25**) (**Fig. 5.4**).

5-3. Synthesis of peptide thioesters

Ac-DYKDDDDKG-COSBn (26) synthesis for test ligations

To establish NCL conditions with the *C*-terminal fragment of the PNT domain (**25**), and also for general use as a tagged-thioester for test ligations with other proteins expressed in the lab, a thioester peptide containing the FLAG sequence (DYKDDDDK) was synthesized. A FLAG-tagged thioester is a useful fragment for test ligations both because it enables detection of the ligation product with high specificity and sensitivity by the FLAG antibody for use in Western blotting, and because it readily dissolves in

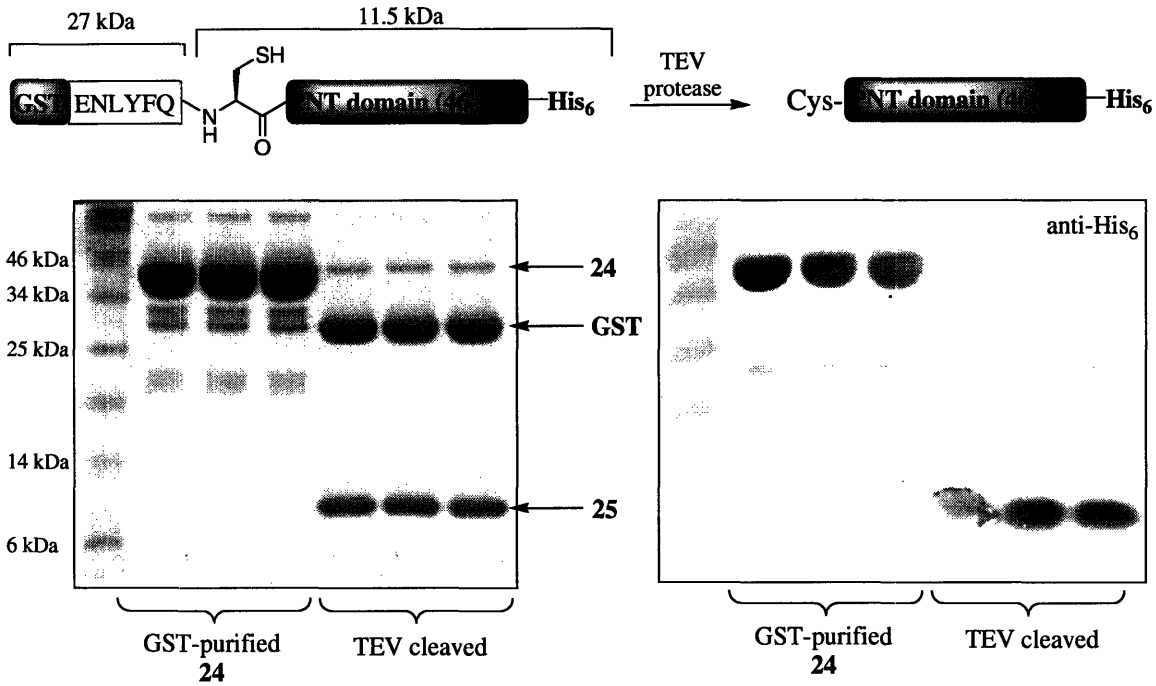


Figure 5.4. Expression and TEV proteolysis of GST-ENLYFQ-Cys-PNT(46-138)-His_6 (39.5 kDa). Coomassie and anti-hexahistidine Western blot analysis of the GST-affinity purified protein **24** and of the product mixture following TEV proteolysis.

standard ligation buffers. For the FLAG thioester synthesis, the peptide was prepared on Gly-loaded TGT resin and cleaved from the resin with 0.5% TFA in DCM to yield a carboxylic acid at the *C*-terminus without effecting side chain protection. Thioesterification was accomplished by activating the carboxylic acid with HBTU and DIPEA and reacting it with benzylmercaptan, as described for paxillin thioesters in Chapter 3 (**Scheme 3.1**). Peptide cleavage and side chain deprotection afforded the FLAG thioester, Ac-DYKDDDDKG-COSBn (**26**). The peptide was purified by reverse phase HPLC, lyophilized, and stored in 400 μg (328 nmol) aliquots at -20 °C for subsequent ligations.

PNTD peptide thioester synthesis

Two peptides were supplied by Elvedin Lukovic on TGT resin for thioesterification and subsequent ligation to **25**: **PNTD-S** (Ac-VPC(Sox)LTPGGRRG-(TGT)); and **PNTD-P**, (Ac-VPC(Sox)LTPGGRRG-(TGT)). The "S" designation refers to non-phosphorylated "substrate" peptide and the "P" designation refers to phosphorylated peptide. The peptide sequences were based on the ERK2

phosphorylation motif of the myelin basic protein (MBP).[12] For peptide thioesterification, the standard conditions used for paxillin and FLAG thioesters were attempted, in which the peptide was activated with HBTU and DIPEA in dry THF and reacted with benzylmercaptan. However, the reaction yielded a single product by HPLC that was 26 to 27 mass units less than that expected for the corresponding PNTD thioester. The identity of this unexpected product could not be determined. While the mass could correspond to hydrolysis of the Sox sulfonamide to sulfonic acid (loss of -NMe_2 and gain of -OH for a net loss of 27 Da), peaks in the MALDI spectrum of the purified peptide suggest that the Sox moiety remains intact. In MALDI spectra of Sox-containing peptides, Sox is often cleaved by the laser, resulting in a second peak at M-264 mass units. If hydrolysis of Sox had occurred, there should have been an M-237 peak in the MALDI, but instead only an M-264 peak was present.

Thioesterification was next attempted using phosphonium reagents in a variety of conditions. These reactions failed to produce the desired thioester and resulted in recovered starting material or in conversion to an unknown product that was 51 mass units smaller than that expected for the thioester. Ultimately activation with HATU and

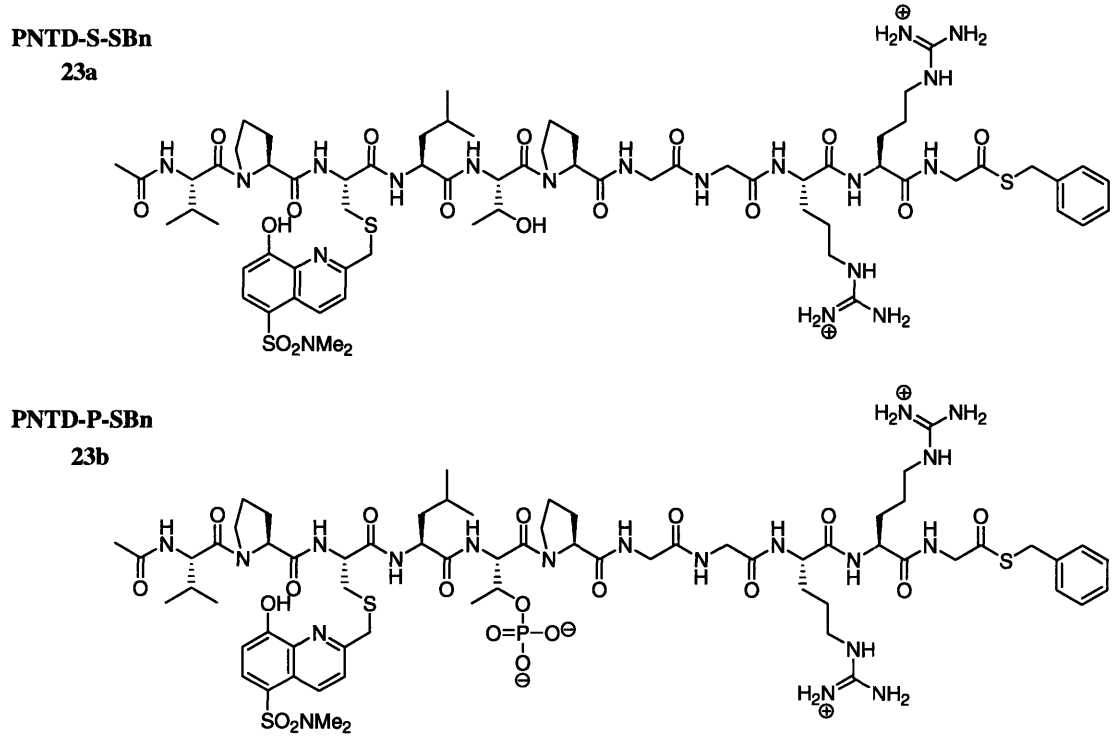


Figure 5.5. PNTD peptide thioesters for ligation to Cys-PNT(46-138)-His_6 (**25**).

HOAT with collidine in DCM resulted in the desired thioester products, **PNTD-S-SBn** (**23a**) and **PNTD-P-SBn** (**23b**) (**Fig. 5.5**). Final side-chain deprotection of the peptide thioesters was tested both in the presence and absence of EDT in the cleavage cocktail, since inclusion of EDT is recommended for cysteine-containing peptides. Both cleavage conditions prevented unwanted cysteine oxidation, so subsequent preparations were conducted in the absence of EDT. The peptide concentrations were determined using the Sox fluorophore (ε_{355} = 8,247 $M^{-1}cm^{-1}$ in 0.1 M NaOH and 1 mM EDTA), and the peptides were stored as lyophilized 100 nmol aliquots at -20 °C for subsequent ligation to Cys-PNT(46-138)-His_6 (**25**).

5.4 Test ligation and product isolation with Cys-PNT(46-138)-His_6

The ligation of the FLAG thioester, **26**, to the PNT fragment **25** was accomplished in a Tris buffer, pH 8, catalyzed by the addition of MESNA. The ligation was run with either 1.5 or 3 equivalents of peptide over protein. With 1.5 equivalents of thioester peptide, the protein conversion to Ac-DYKDDDDKGC-PNT(54-138)-His_6 (**27**), as estimated by Coomassie gel analysis, was 60%. With 3 equivalents of thioester the protein conversion was approximately 90%.

The semisynthesis was designed to enable complete purification of the ligation product by buffer exchange and a single affinity column, based on the *N*-terminal FLAG tag. This means that a separate purification step was not required following the TEV protease cleavage prior to ligation. The crude ligation mixture therefore contained cleaved GST tag (27 kDa), hexahistidine-tagged TEV protease (~ 28 kDa), NCL product (**27**) (12.7 kDa), unligated PNT domain (**25**) (11.5 kDa), unligated thioester (**26**) (1.2 kDa) and MESNA (**Fig. 5.6a**). Exchange into TBS with 10 kDa MWCO centrifugal filters concurrently removed the MESNA and excess thioester. Incubation with FLAG-affinity resin isolated **27** as the sole FLAG-tagged moiety in the NCL mixture. Coomassie and anti-hexahistidine Western blot analysis revealed complete separation of ligated product from the unligated material, as well as excellent separation from cleaved GST (**Fig. 5.6b**). This purification strategy can be applied to ERK2 sensor isolation if an *N*-terminal purification handle is included on the peptide thioester. In that case, the ligation would be run with 1.5 equivalents of thioester to minimize the use of excess Sox-

containing peptides. However, the first generation sensor does not include an *N*-terminal handle, since the FLAG sequence was found to decrease chelation-enhanced fluorescence of test peptides. Purification is more complicated in the absence of an *N*-terminal tag, so 6 equivalents of thioester over PNT domain were used to increase product conversion for the ERK2 sensor, as described in Section 5.5.

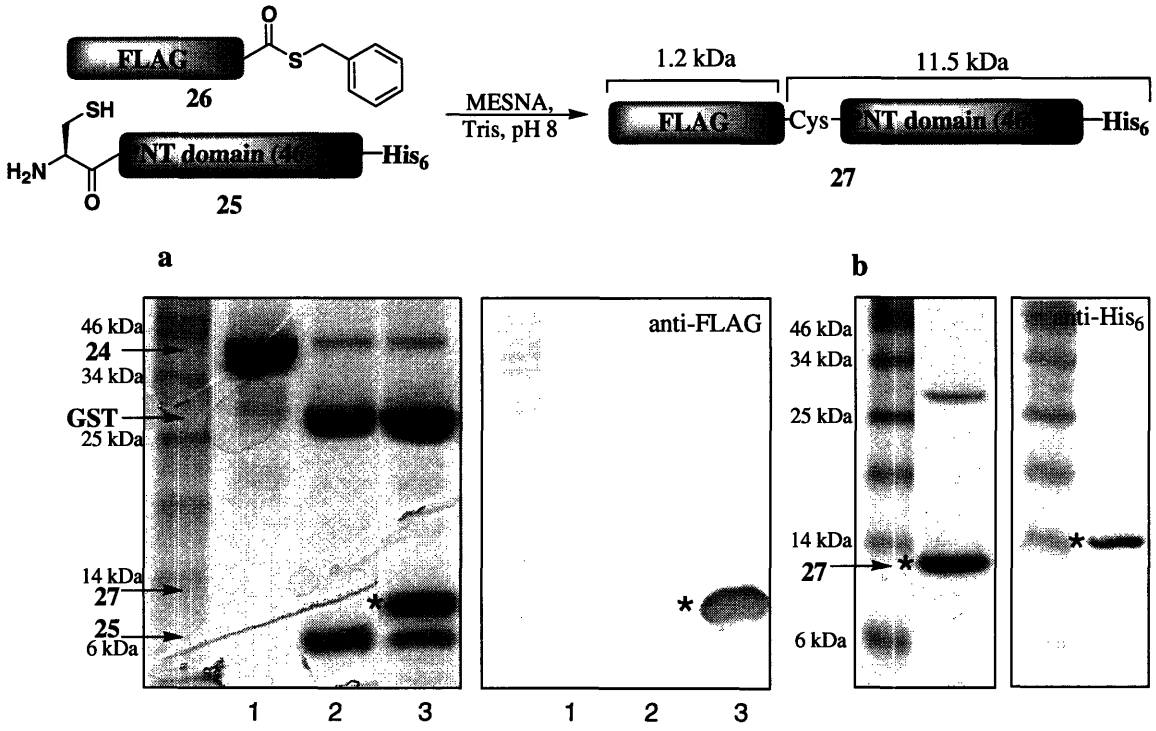


Figure 5.6. Test ligation and purification of Ac-FLAG-Cys-PNT(46-138)-His_6 (**27**). **a**) Coomassie and anti-FLAG Western blot analysis of (lane 1) the purified GST-fusion construct **24**, (lane 2) the TEV cleavage product **25**, and (lane 3) the crude NCL product mixture. Only the ligated product should be visible on an anti-FLAG Western blot. **b**) Coomassie and anti-hexahistidine Western blot analysis of FLAG-affinity purified **27**, The red stars indicate the desired ligation product, **27**.

Stability and buffer compatibility of the ligated PNT domain

GST-ENLYFQ-Cys-PNT(46-138)-His_6 (**24**) is very stable prior to cleavage of the GST tag and may be stored at 0.5 mg/mL for over one month at 4 °C without any detectible precipitation or degradation. Following GST cleavage and concentration (to 3 mg/mL) of the resulting protein construct **25**, some precipitation was observed, which could be avoided by immediate ligation with an *N*-terminal fragment. To probe the

stability of the ligated PNT domain, Ac-DYKDDDDKGC-PNT(46-138)-His_6 (**27**) was dialyzed into four different buffers at 0.1 mM immediately following NCL and stored for one week at 4 °C. The protein was also analyzed in the crude ligation mixture at 0.25 mM. Stability was examined in TBS, PBS, 20 mM hepes, pH 7.4 and 20 mM hepes, pH 7.4 with 100 mM NaCl, the standard buffer for Sox fluorescence assays. The protein was found to be stable and free of precipitation in all conditions tested (**Fig. 5.7**).

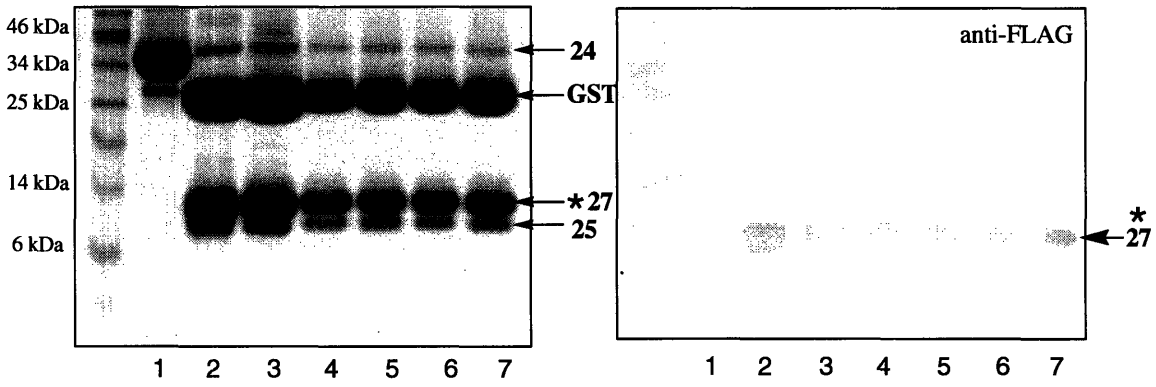


Figure 5.7. Coomassie and anti-FLAG probed Western blot analysis of PNT stability following NCL. (lane 1) purified GST-fusion construct **24,** (lane 2) crude NCL product, and (lane 3-7) NCL product after 1 week at 4 °C in (lane 3) ligation buffer (2.5 mg/mL), (lane 4) TBS (1 mg/mL), (lane 5) PBS (1mg/mL), (lane 6) hepes (1 mg/mL), and (lane 7) hepes with 100 mM NaCl.

5.5 NCL of the ERK2 kinase sensor (22a)

Following the validation of ligation conditions and product stability using the FLAG thioester **26, PNTD-S-SBn (23a)** was ligated to Cys-PNT(54-138)-His_6 (**25**) in Tris buffer, pH 8 with 150 mM MESNA. The peptide and PNT domain concentrations were 0.8 mM and 0.25 mM, respectively, and the conversion to ligation product, Ac-VPC(Sox)LTPGGRRG-PNT(54-138)-His_6 (**22a**), was approximately 70%. Increasing the thioester excess from 3 to 6 equivalents resulted in over 90% conversion (**Fig. 5.8**). Isolation of **22a** (13 kDa), from unligated thioester (~ 1.5 kDa) (**23a**) was performed by Elvedin Lukovic using size exclusion chromatography. The phosphorylated version of the sensor (**22b**) was constructed using the identical ligation conditions to conjugate **25** to **PNTD-P-SBn (23b)**.

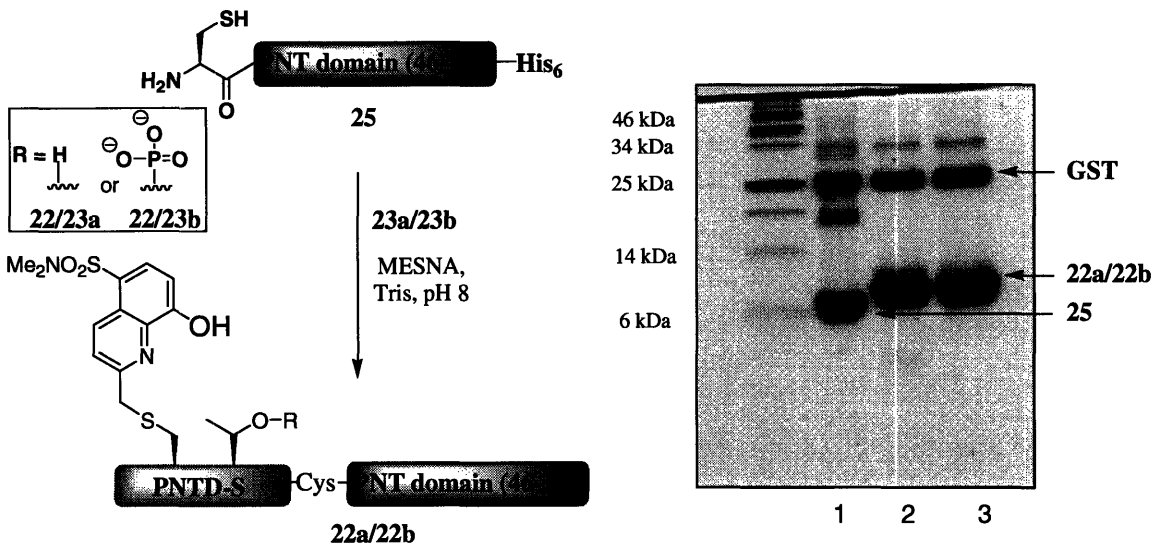


Figure 5.8. NCL to generate Ac-VPC(Sox)LTPGGRRG-Cys-PNT(46-138)-His_6 (**22a**) and Ac-VPC(Sox)LpTPGGRRG-Cys-PNT(46-138)-His_6 (**22b**). Coomassie gel of (lane 1) the TEV cleavage product **25** (11.5 kDa), (lane 2) the crude NCL product **22a** (13 kDa), and (lane 3) the crude NCL product **22b** (13 kDa).

Fluorescence assays performed by Elvedin Lukovic revealed a three-fold increase in fluorescence of the phosphorylated sensor (**22b**) over the substrate sensor (**22a**). In addition, the ERK2 sensor (**22a**) demonstrated excellent sensitivity to ERK2 activity. Under identical assay conditions, the protein sensor was found to be approximately 10-fold more sensitive than the corresponding peptide probe, **23a**, for monitoring ERK2 kinase activity (**Fig. 5.9**).

Conclusion

The semisynthesis of the ERK2 kinase activity sensor described in this chapter extends the Sox-sensing strategy to an 11.5-kDa protein domain for the creation of a sensor that comprises both an ERK2 phosphorylation motif and the Ets-1 PNT domain for ERK2 docking. The details of sensor construction were based on the successful use of tags and a TEV cleavage site for the semisynthesis of paxillin analogs (Chapter 3). Results from fluorescence assays on the completed sensor indicate that the PNT domain confers sensitivity for ERK2 activity that was not possible with a linear peptide-probe for ERK2 phosphorylation. We have therefore demonstrated the use of semisynthesis to

enhance the sensitivity of kinase sensors by incorporating extended recognition elements onto the peptide recognition core.

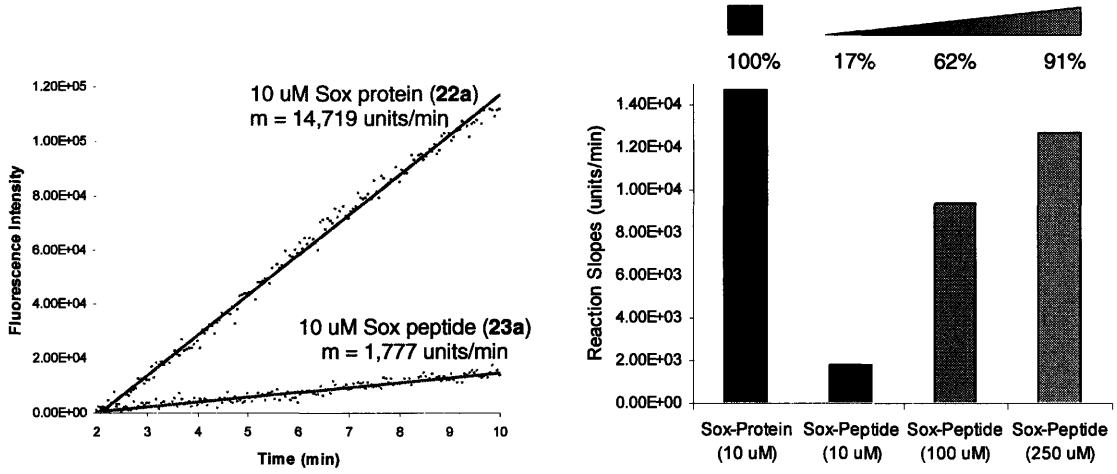
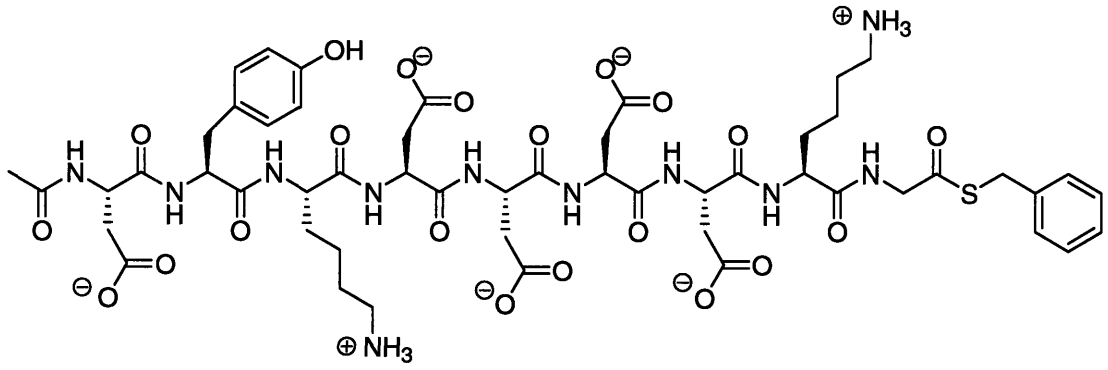


Figure 5.9. (Assays performed by Elvedin Lukovic.) Direct comparison of the Sox-protein (**22a**) versus the Sox-peptide (**23a**) as sensors for ERK2 phosphorylation. The graph on the left shows the fluorescence intensity of the probes over time following treatment with ERK2. The bar graph on the right demonstrates the percentage of the Sox-peptide slopes with respect to the protein slope with 10, 100 and 250 μM of peptide sensor (**23a**).

Methods

Ac-DYKDDDDKG-COSBn (26).



The peptide (50 μmol on 250 mg of resin) was synthesized on an automated Advanced ChemTech 396 peptide synthesizer using standard Fmoc-based SPPS protocols on preloaded Fmoc-Gly-Novasyn TGT resin (Novabiochem). The peptide was cleaved from the resin with side-chain protection intact by agitating with 0.5% TFA in

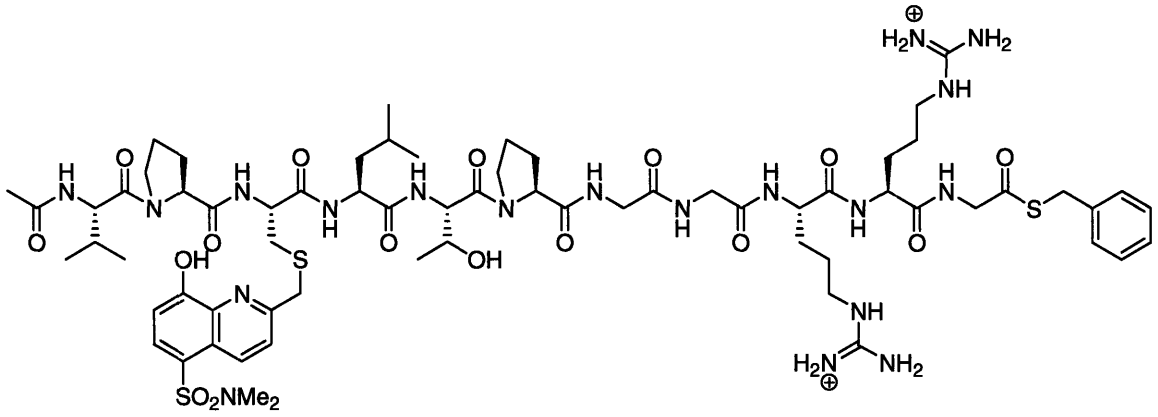
DCM for 1.5 h. The resin was removed by filtration and rinsed with DCM. The solvent was evaporated under a stream of nitrogen, and the peptide was triturated with cold hexanes. The hexanes were removed in vacuo and the resulting white powder was dissolved in anhydrous THF (17 mL), then treated with HBTU (75 mg, 200 μmol), DIPEA (69 μL, 400 μmol), and benzylmercaptan (23 μL, 200 μmol) under nitrogen overnight. The THF was removed in vacuo and the peptide was deprotected with 95% (vol/vol) TFA, 2.5% (vol/vol) TIS, and 2.5% (vol/vol) water for 2 hours. The peptide was triturated with cold diethylether, and purified by reverse phase HPLC on a Waters 600 instrument with a YMC C_{18} preparative column using an elution gradient of water/acetonitrile with 0.1% TFA. The identities of the peptide as a free acid and of the final peptide thioester product were confirmed by electrospray ionization (ESI) mass spectrometry on a Perspective Biosystems *Mariner Biospectrometry Workstation* (turbo ion source).

Ac-DYKDDDDKG-COOH, Reverse phase HPLC (t_R = 18.1 min). Exact mass calcd for $C_{45}H_{65}N_{11}O_{22}$, 1112.1; found by MS(ESI), 1112.3 $[MH]^{1+}$, 556.7 $[MH_2]^{2+}$.

Ac-DYKDDDDKG-COSBn (**26**), Reverse phase HPLC (t_R = 21.3 min). Exact mass calcd for $C_{52}H_{71}N_{11}O_{21}S$, 1217.8; found by MS(ESI), 1218.4 $[MH]^{1+}$, 609.7 $[MH_2]^{2+}$.

Sox-containing peptides thioesters. Peptides were constructed by Fmoc-based SPPS on Fmoc-Gly-Novasyn TGT resin and derivatized to the corresponding thioesters as described below for **PNTD-S-SBn**.

PNTD-S-SBn *(Ac-VPC(Sox)LTPGGRRG-COSBn) (**23a**).*

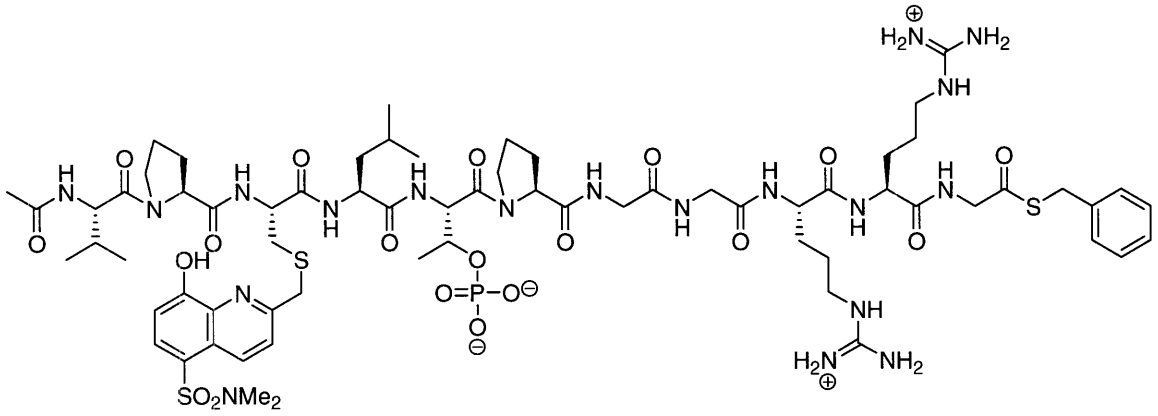


Following SPPS on Fmoc-Gly-Novasyn TGT resin, the peptide (10 μmol from 50 mg of resin) was cleaved from the resin with side-chain protection intact by agitating with 0.5% TFA in DCM for 1.5 h. The resin was removed by filtration and rinsed with DCM, the solvent was mostly evaporated under a stream of nitrogen, and the peptide was triturated with cold hexanes. The hexanes were removed in vacuo, and the resulting white powder was dissolved in 16 mL of anhydrous DCM and treated with HATU (30 mg, 80 μmol) and HOAT (10.8 mg, 80 μmol) dissolved in 0.5 mL dry DMF, collidine (21 μL, 160 μmol), and benzylmercaptan (19 μL, 160 μmol). The reaction was stirred under nitrogen overnight. The solvent was removed in vacuo, and the peptide was deprotected with 95% (vol/vol) TFA, 2.5% (vol/vol) TIS, and 2.5% (vol/vol) water for 3 h. The peptide was triturated with cold diethylether, and purified by reverse phase HPLC.

PNTD-S: Ac-VPC(Sox)LTPGGRRG-COOH, Reverse phase HPLC (t_R = 25.8 min). Exact mass calcd for $C_{60}H_{95}N_{19}O_{17}S_2$, 1417.7; found by MS(MALDI), 1419.3 $[MH]^{1+}$.

PNTD-S-SBn: Ac- VPC(Sox)LTPGGRRG –COSBn (**23a**), Reverse phase HPLC (t_R = 27.2 min). Exact mass calcd for $C_{67}H_{101}N_{19}O_{16}S_3$, 1523.7; found by MS(MALDI), 1524.8 $[MH]^{1+}$.

***PNTD-P-SBn**: Ac- VPC(Sox)LpTPGGRRG –COSBn (**23b**).*



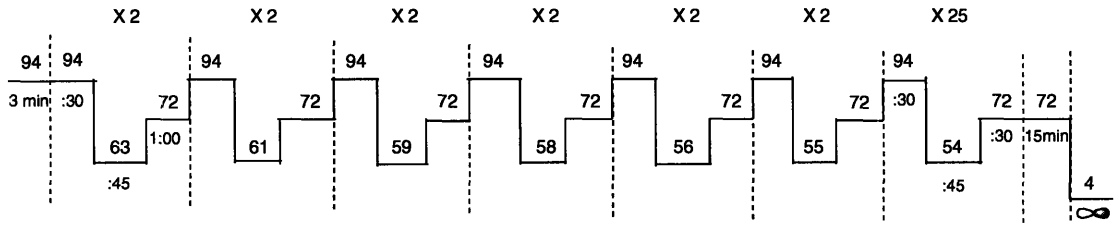
PNTD-P: Ac-VPC(Sox)LpTPGGRRG-COOH, Reverse phase HPLC (t_R = 25.8 min). Exact mass calcd for $C_{60}H_{96}N_{19}O_{20}PS_2$, 1497.6; found by MS(MALDI), 1498.4 $[MH]^{1+}$.

PNTD-P-SBn: Ac- VPC(Sox)LpTPGGRRG –COSBn (**23b**), Reverse phase HPLC (t_R = 27.2 min). Exact mass calcd for $C_{67}H_{102}N_{19}O_{19}PS_3$, 1603.7; found by MS(MALDI), 1605.9 $[MH]^{1+}$.

*Plasmid construction for GST-ENLYFQC-PNT(46-138)-His_6 (**24**).* A kanamycin-resistant Ets S26A construct in a pET28 vector was transformed into DH5α cells and grown up on an LB-kan agar plate. Several colonies were selected and grown up in 3 mL of LB-carbenicillin overnight with shaking at 37 °C. The plasmids from these cultures were isolated using a perfect prep plasmid mini kit (Eppendorf). PNT(46-138) was amplified from the plasmid DNA from Ets miniprep #4 using primers to introduce a 5' EcoR1 site and a 3' Not1 site for incorporation into a pGEX-4T-2 vector. The primers also encoded an amino-terminal TEV protease recognition site (ENLYGQ✂C) and a carboxy-terminal hexahistidine tag. For this amplification the following primers were used: **FORETS1**: 5'- GCC GGA ATT CGT GAA AAC CTG TAT TTT CAG TGC TCC CAA GCC TTG AAA GCT -3;, and **BACETS1**: 3'- GCC CCC TTT TGC GGC CGC CTA GTG ATG GTG ATG GTG ATG ACC TTT CAC ATC CTC TTT CTG -5'.

Five PCR reactions were run simultaneously with 50 µL each of the following reaction mixture: 200 µL of sterile DI water, 25 µL of 10X Taq HIFI buffer, 16.25 µL of 10 mM dNTPs, 9.75 µL of 50 mM $MgCl_2$, 3.25 µL (0.5 mg/mL) each of FORETS1 and BACETS1 primers, 3.5 µL of ETS1 miniprep 4 DNA template, and 2.5 µL of Taq HIFI

(added last). The reactions were run in the thermal cycler as shown and electrophoresed on an agarose gel. The desired DNA fragment (354 bp) was excised and purified with a gel extraction kit (Qiagen).



The PCR-amplified fragments were digested with Not1 and EcoR1 and ligated to Not1/EcoR1-digested and CIP-treated pGEX-4T-2. To 43.5 μL of purified PCR product in EB were added 5.0 μL of 10X EcoR1 buffer, 0.5 μL of 100X BSA, and 2.5 μL of Not1. The digestion was incubated overnight at rt, and then 2.5 μL of EcoR1 were added, and the mixture was further incubated for 1 hour at 37 °C. To this solution was added 12.5 μL of 5X nucleic acid loading buffer and the digested insert was electrophoresed on a 1% agarose DNA gel for purification. The insert band was excised and purified with a gel extraction kit (Qiagen). The DNA was eluted with 30 μL of 1:9 elution buffer: sterile water and used directly in ligations without any further concentration.

To ligate the PNT(46-138)-His_6 insert into the pGEX-4T-2 vector, 3 μL of digested and CIP-treated vector, 5 μL of digested (334 bp) PCR insert, 1 μL of T4 ligase, and 1 μL of 10X T4 ligase buffer were combined in a sterile eppendorf. The tube was gently flicked and then incubated 2.5 h at rt. A vector-only control ligation was similarly run. The ligation mixture was transformed into DH5α cells and grown on carbenicillin-resistant plates. Eight colonies grew up on the vector/insert ligation plate with only two colonies on the negative control. Plasmid DNA was isolated from selected colonies and confirmed by sequencing. Miniprep #2 was found to contain the correct insert.

Expression and purification of GST-ENLYFQC-PNT(46-138)-His_6 (24). The PNT domain plasmid was transformed into BL21(DE3) competent cells (Stratagene) and grown at 37 °C to midlog phase (OD = 0.55) in 1 L of LB media with carbenicillin. The culture was cooled to 30 °C, and the cells were induced with 0.2 mM IPTG and grown

for an additional 4.5 h. Cells were harvested by centrifugation and frozen at -80 °C. For cell lysis, the pellet from a 1-L expression was thawed and resuspended in 40 mL of lysis buffer (PBS, 1 mg/mL lysozyme, 1 mM DTT, and protease cocktail III [100 μM AEBSF, 80 nM aprotinin, 5 μM bestatin, 1.5 μM E-64, 2 μM leupeptin, 1 μM pepstatin A] (Calbiochem)) and incubated for 20 min at 4 °C. The cells were lysed with 1% NP-40 Alternative, then sonicated and subjected to centrifugation at 9,000 rpm for 25 min, followed by filtration through a 0.2 micron filter. The supernatant was transferred to a 50-mL conical tube for purification.

The soluble fraction was incubated with 3 mL of Glutathione Sepharose 4 Fast Flow resin (Amersham Biosciences), following the manufacturer protocol. After 35 min of incubation at 4 °C and copious washing with PBS, the protein was eluted from the resin with four 3-mL batches of 10 mM glutathione in 50 mM Tris, pH 8.0, following incubation for 3-5 min. The elutions were combined and dialyzed into 3 X 2 L of PBS to remove the glutathione. A 1-L prep afforded 25 to 30 mg of GST-ENLYFQC-PNT(46-138)-His_6. The 38.5-kDa protein was analyzed by 10% SDS-PAGE gels and visualized with Coomassie blue dye and by Western blot with a mouse anti-hexahistidine primary antibody. Purified **24** was stored at 4 °C.

*TEV proteolysis. Cys-PNT(46-138)-His_6 (**25**):* Purified GST-ENLYFQC-PNT(46-138)-His_6 was diluted to 1 mg/mL into a TEV cleavage buffer with a final concentration of 50 mM Tris pH 8.0, 500 μM EDTA, and 5 mM BME. TEV protease (43 μL of protease per mg of protein) was added, and the resulting solution was incubated at 27 °C for 3 hours. The protein was concentrated in a swinging bucket centrifuge at 2 K rpm using 5-KDa MWCO Amicon Ultra-15 centrifugal filters (Millipore) and used immediately in NCL without removal of the TEV or the cleaved GST tag. Cleavage of 10 mg of GST-ENLYFQC-PNT(46-138)-His_6 resulted in 3 mg of Cys-PNT(46-138)-His_6 with complete conversion.

Ligations

NCL was carried out with 250 μM protein, 0.8 mM peptide thioester, and 150 mM MESNA in TBS at pH 8.0, unless noted otherwise in the text. For example, to a

solution of Cys-PNT(46-138)-His_6 (104 μg, 104 nmol) in 330 μL of TBS was added **PNTD-S-SBn** (lyophilized, then dissolved into 20 μL of water for transfer) (327 nmol), 30 μL of 2 M MESNA, and 30 μL of 500 mM Tris pH 8.0. The reaction was incubated for 20 h at 25 °C, and then dialyzed into PBS using a 10 MWCO dialysis membrane to remove excess peptide thioester. The protein was analyzed by 10% SDS-PAGE gels, and visualized with Coomassie blue dye.

Acknowledgements

I thank Elvedin Lukovic for being a fantastic partner on this project and for his expertise on the Sox amino acid. As mentioned in the text, Elvedin supplied the PNTD peptides on TGT resin for thioesterification. Elvedin also carried out the final sensor purification and is currently performing experiments on the purified sensors. We thank Professor Kevin Dalby for donating the Ets S26A construct.

References

1. Blume-Jensen, P. & Hunter, T. Oncogenic kinase signaling. *Nature (London, United Kingdom)* **411**, 355-365 (2001).
2. Shults, M. D., Carrico-Moniz, D. & Imperiali, B. Optimal Sox-based fluorescent chemosensor design for serine/threonine protein kinases. *Anal. Biochem.* **352**, 198-207 (2006).
3. Shults, M. D., Pearce, D. A. & Imperiali, B. Modular and Tunable Chemosensor Scaffold for Divalent Zinc. *J. Am. Chem. Soc.* **125**, 10591-10597 (2003).
4. Shults, M. D. & Imperiali, B. Versatile Fluorescence Probes of Protein Kinase Activity. *J. Am. Chem. Soc.* **125**, 14248-14249 (2003).
5. Shults, M. D., Janes, K. A., Lauffenburger, D. A. & Imperiali, B. A multiplexed homogeneous fluorescence-based assay for protein kinase activity in cell lysates. *Nat. Methods* **2**, 277-283 (2005).
6. Walsh, C. T. *Posttranslational modification of proteins: expanding nature's inventory* (Roberts and Company, Englewood, 2006).

7. Foulds, C. E., Nelson, M. L., Blaszczak, A. G. & Graves, B. J. Ras/mitogen-activated protein kinase signaling activates Ets-1 and Ets-2 by CBP/p300 recruitment. *Mol. Cell. Biol.* **24**, 10954-10964 (2004).
8. Seidel, J. J. & Graves, B. J. An ERK2 docking site in the Pointed domain distinguishes a subset of ETS transcription factors. *Genes Dev.* **16**, 127-137 (2002).
9. Rainey, M. A., Callaway, K., Barnes, R., Wilson, B. & Dalby, K. N. Proximity-Induced Catalysis by the Protein Kinase ERK2. *J. Am. Chem. Soc.* **127**, 10494-10495 (2005).
10. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science (Washington, D. C.)* **266**, 776-9 (1994).
11. Hackeng, T. M., Griffin, J. H. & Dawson, P. E. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10068-10073 (1999).
12. Haycock John, W. Peptide substrates for ERK1/2: structure-function studies of serine 31 in tyrosine hydroxylase. *J. Neurosci. Meth.* **116**, 29-34 (2002).

Appendix

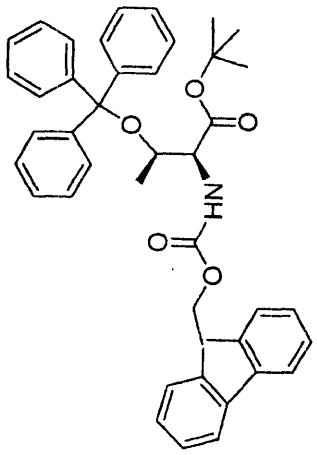
NMR Spectra

1H NMR of N-α-Fmoc-hydroxyltrityl-L-threonine tert-butyl ester (14)

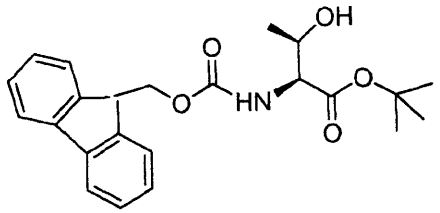
N-α-Fmoc-hydroxytrityl-L-Threonine tert-butyl ester

13C NMR of N-α-Fmoc-hydroxyltrityl-L-threonine tert-butyl ester (14)

N-α-Fmoc-hydroxytrityl-L-Threonine tert-butyl ester

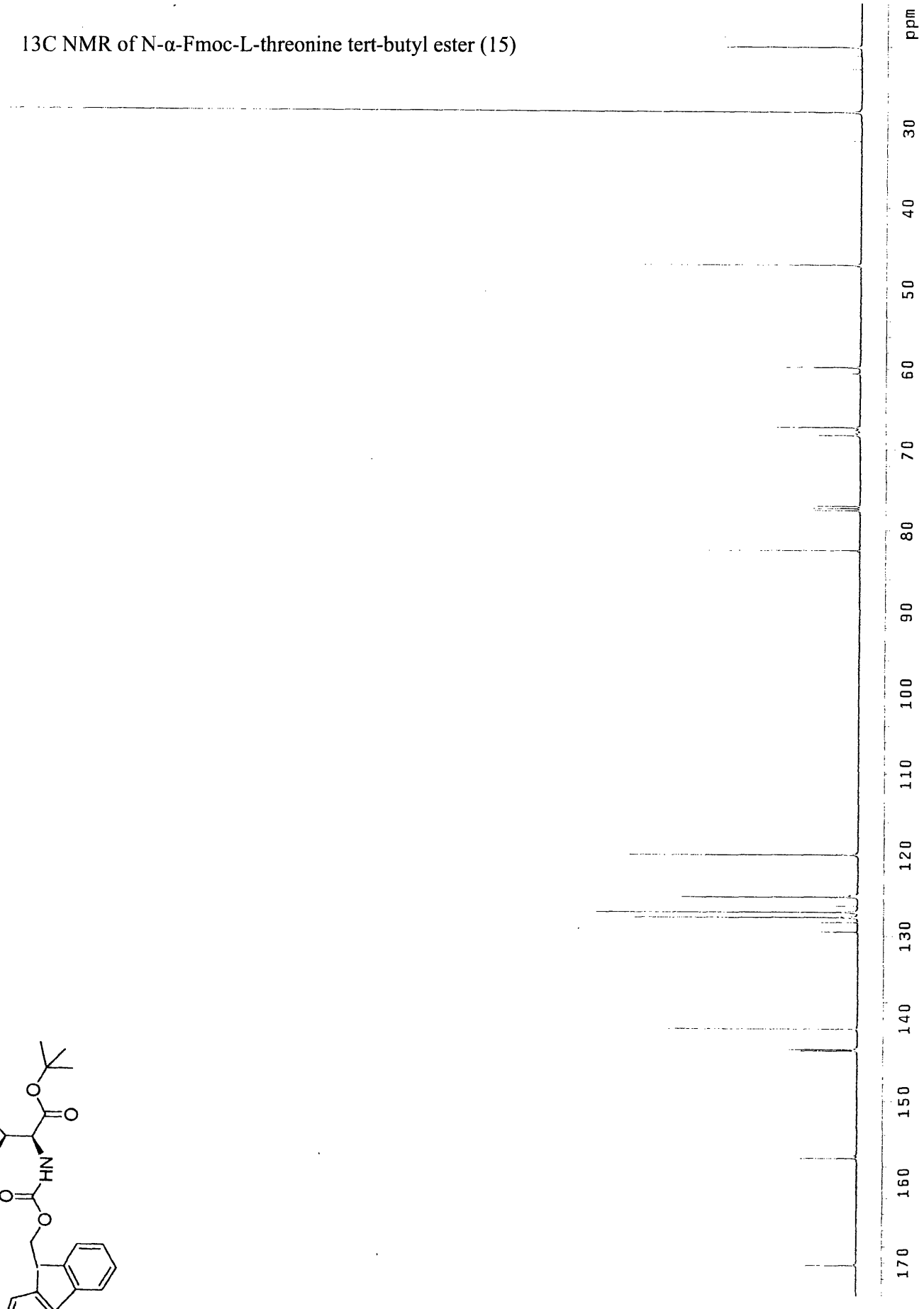


N-α-Fmoc-L-Threonine tert-butyl ester



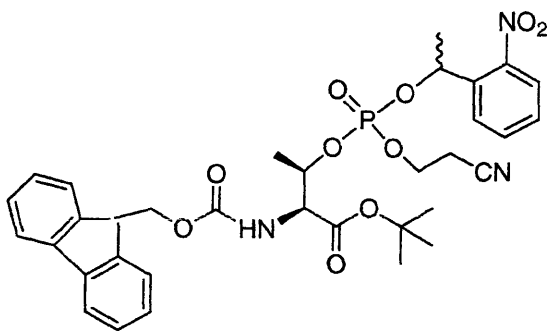
1H NMR of N-α-Fmoc-L-threonine tert-butyl ester (15)

13C NMR of N-α-Fmoc-L-threonine tert-butyl ester (15)



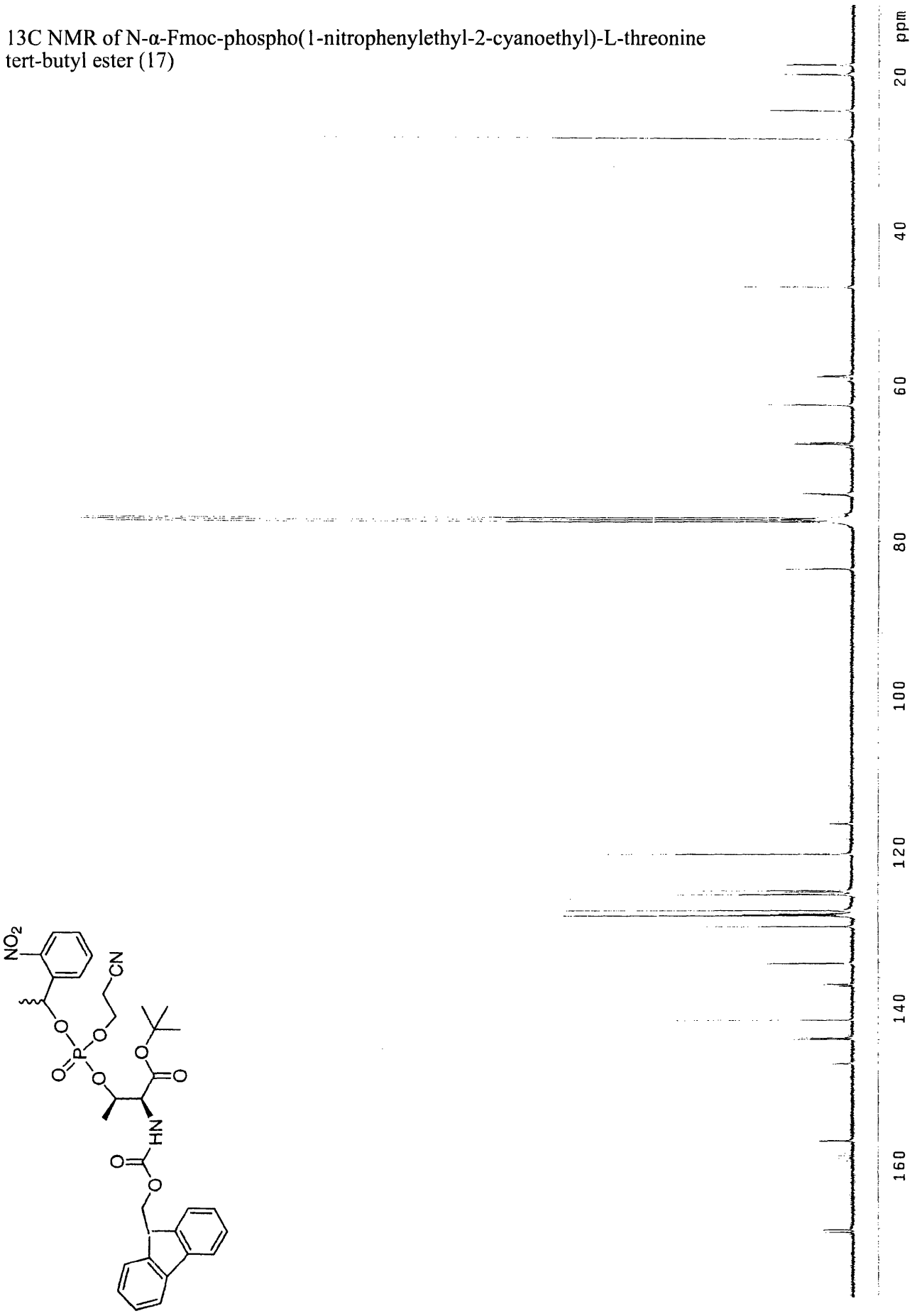
N-α-Fmoc-L-Threonine tert-butyl ester

N-α-Fmoc-phospho-(1-nitrophenylethyl-2-cyanoethyl)-L-Threonine tert-butyl ester



1H NMR of N-α-Fmoc-phospho(1-nitrophenylethyl-2-cyanoethyl)-L-threonine tert-butyl ester (17)

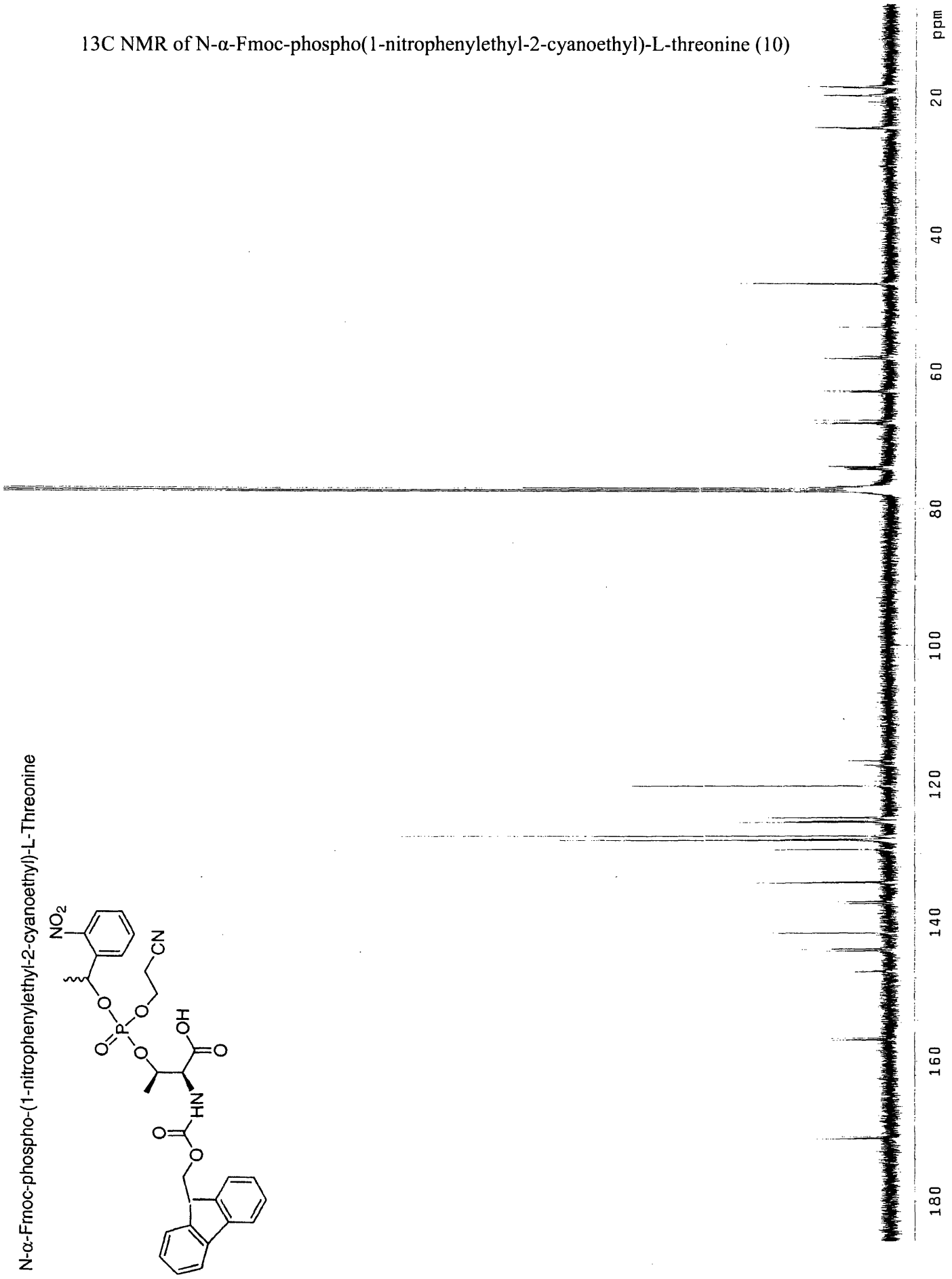
13C NMR of N-α-Fmoc-phospho(1-nitrophenylethyl-2-cyanoethyl)-L-threonine tert-butyl ester (17)



N-α-Fmoc-phospho-(1-nitrophenylethyl-2-cyanoethyl)-L-Threonine tert-butyl ester

1H NMR of N-α-Fmoc-phospho(1-nitrophenylethyl-2-cyanoethyl)-L-threonine (10)

13C NMR of N-α-Fmoc-phospho(1-nitrophenylethyl-2-cyanoethyl)-L-threonine (10)



N-α-Fmoc-phospho-(1-nitrophenylethyl-2-cyanoethyl)-L-Threonine

Elizabeth M. Vogel

Department of Chemistry
MIT
400 Main Street
Cambridge, MA 02139
tel. 617-253-0206 (lab)
evogel@mit.edu

EDUCATION/ EXPERIENCE

September 2001– Present — **Massachusetts Institute of Technology** — **Cambridge, Mass.**
Ph.D. program, Department of Chemistry
Graduate student of Professor Barbara Imperiali
GPA: 5.0/ 5.0

August 1997- May 2001 — **Hamilton College** — **Clinton, New York**
B.A. (Chemistry, ACS), *summa cum laude*
Valedictorian
English Minor, GPA: 4.0/4.0
Undergraduate research with Professor Ian J. Rosenstein

TEACHING

2003 Organic Chemistry I, head teaching assistant – MIT
2002 Intermediate Chemical Experimentation, teaching assistant – MIT
2001 Introductory Chemical Experimentation, teaching assistant - MIT
1999-2001 Organic chemistry tutor - Hamilton College

HONORS

Graduate

Pfizer Poster Award
- *2006 Gordon Research Conference on Bioorganic Chemistry*

Charles Krakauer Graduate Fellowship
- *2005-2006, MIT Center for Cancer Research*

Wyeth Scholarship
- *2005, awarded in recognition of research, based on 4th year research seminar*

ACS/ GDCh Exchange Travel Award
- *2005*

Merck/MIT Computational and Systems Biology Fellowship
- *2003-2005*

MIT Chemistry Department Award for Outstanding Teaching by a Graduate Student
- *2002*

Undergraduate

Valedictorian
- *2001, Hamilton College*

Phi Beta Kappa

Norton Prize in Chemistry
- *2001, awarded to the undergraduate who demonstrates the greatest capacity for research*

Underwood Prize in Chemistry
- *2001, awarded to a senior who excels in chemistry*

Donald J. Denny Prize in Physical Chemistry
- *2001*

Tarbell Book Prize in Organic Chemistry
- *1999*

PUBLICATIONS

Vogel, E. M., Imperiali, B. Native chemical ligation (NCL): Semisynthesis of post-translationally modified proteins and biological probes. In: RajBhandary, T., Koehrer, C. (ed) Protein Engineering, vol 17. Springer, Berlin Heidelberg New York. *Invited book chapter, in press.*

Vogel, E. M., Imperiali, B. Semisynthesis of unnatural amino acid mutants of paxillin: Protein probes for cell migration studies. *Protein Sci.* **2007**, *in press.*

Rothman, D. M., Vasquez, M. E., **Vogel, E. M.**, Imperiali, B. Caged Phospho-Amino Acid Building Blocks for Solid Phase Peptide Synthesis. *J. Org. Chem.* **2003**, *68(17)*, 6795-6798.

Rothman, D. M., Vasquez, M. E., **Vogel, E. M.**, Imperiali, B. General Method for the Synthesis of Caged Phosphopeptides: Tools for the Exploration of Signal Transduction Pathways. *Org. Lett.* **2002**, *4*, 2865-2868.

POSTERS

Caged Phosphorylated Paxillin: Semisynthesis and Application of Paxillin Analogs for Cell Migration Studies. Vogel, E. M., Imperiali, B. *Gordon Research Conference on Bioorganic Chemistry, July 2006, Oxford, England.*

Caged Phosphoproteins: Semisynthesis of Caged PhosphoPaxillin. Vogel, E. M., Imperiali, B. *Euregionale Student Research Conference, April 2005, Berlin, Germany.*

Probing for Electronic Effects in the Transition State of the Cyclopropylcarbinyl Radical Ring Opening Reaction. Vogel, E. M., Rosenstein, I. J. *37th National Organic Chemistry Symposium, 2001, Bozeman, Montana.*